

Open Research Online

The Open University's repository of research publications and other research outputs

Children making faces: enhancing children's facial recall and composite construction

Thesis

How to cite:

Paine, Carina Bridget (2005). Children making faces: enhancing children's facial recall and composite construction. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2005 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.21954/ou.ro.0000e949>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk



**CHILDREN MAKING FACES:
ENHANCING CHILDREN'S FACIAL RECALL
AND COMPOSITE CONSTRUCTION**

Thesis submitted for the Degree of Doctor of Philosophy

by

Carina Bridget Paine

B.Sc (Hons).

Department of Psychology

Faculty of Social Sciences

The Open University, UK

2004

Volume 1 of 2

Author No: TS389382
Submission date: 23 November 2004
Award date: 8 April 2005

ABSTRACT

When the identity of an offender is unknown in a criminal investigation witnesses are often asked to produce a verbal description and/or a facial ‘composite’ image of the suspect’s face with a police operator. The aims of the research presented in this thesis were (1) to gain an understanding of children’s verbal descriptions of unfamiliar faces, and (2) to explore how children might be assisted through appropriate interview techniques to provide computerised facial composite constructions of unfamiliar faces. The thesis involved an innovative synthesis of developmental, eyewitness, facial recognition and facial recall research in order to address these aims. The thesis comprised four studies as follows:

Study 1 involved a questionnaire survey of facial composite operators in the UK to identify their current practice with, and experiences and opinions of, child witnesses. Analyses of the data indicated two main issues facing operators when interviewing child witnesses: (1) children’s language (problems describing and/or understanding); and (2) children’s lack of concentration. The results of this survey and existing research informed a series of laboratory-based experimental studies with children aged 6-, 8- and 10-years.

Study 2 (Experiments 1 and 2) explored, for the first time in the existing literature, the language and terms children use to describe unfamiliar faces. The findings showed that from the age of 6-years, children were able to effectively describe unfamiliar faces when prompted (in terms of the content and quantity of descriptions useful for subsequent composite construction). The descriptions provided were used to produce original sets of developmentally appropriate verbal and visual prompts.

Study 3 (Experiments 3 and 4) investigated the effect of the verbal and visual prompts as potentially appropriate interview techniques on the production of children's descriptions of unfamiliar faces. Results suggested that the visual prompts significantly enhanced children's descriptions of unfamiliar faces, in terms of increasing the number and accuracy of descriptions, and reducing the time taken.

Study 4 (Experiments 5 through 7) explored the use of the prompts as potentially appropriate interview techniques to assist children's composite constructions of unfamiliar faces with the computerised composite system E-FIT. The quality of the composites constructed was evaluated using subjective likeness rankings and ratings, as well as an objective utility measure. Results demonstrated that children from the age of 6-years were able to produce facial composites of an unfamiliar face. Although results showed that composite construction drew on developmentally sensitive skills, there were large variations within age groups and composites produced by adults and older children were not always evaluated as better than younger children's.

Collectively the findings from the thesis are encouraging, indicating that children should not be excluded from providing facial descriptions and constructing facial composites simply on the basis of their age.

Finally, the theoretical contributions and practical implications of the findings are discussed.

DECLARATION

This thesis consists of my own original work and comprises less than 100,000 words (inclusive of tables, references, and appendices). Due acknowledgement has been made to all material used within this thesis where appropriate.

Financial support was provided by a doctoral research studentship from The Open University.

The research has been presented at a number of conferences:

Paine, C.B., Brace, N.A., Pike, G.E., and Westcott, H.L. (2002)

Paine, C.B., Brace, N.A., Pike, G.E., and Westcott, H.L. (2003a)

Paine, C.B., Brace, N.A., Pike, G.E., and Westcott, H.L. (2003b)

Paine, C.B., Brace, N.A., Pike, G.E., and Westcott, H.L. (2003c)

ACKNOWLEDGMENTS

Foremost, I would like to offer my very sincere thanks to my supervisors, Dr. Helen Westcott, Dr. Nicola Brace and Dr. Graham Pike. This thesis is a result of their continuous support, encouragement, and guidance every step of the way.

I would also like to thank some of the people outside of my supervisory team who have assisted me in the completion of this thesis in various ways:

I want to express my appreciation to everyone at the schools where I conducted my research. Thank you to the head teachers, class teachers and parents who made me extremely welcome and went out of their way to assist me with my research. I would especially like to thank Lisa for her help and support and of course, all of the children who participated in the studies and pilot studies.

Thanks must also go to various expert police personnel who willingly shared their invaluable knowledge with me. To Jan Richardson, Peter Bennett, Cathy Charsley, Chris Few, and in particular Anne Parry who quickly became, and I hope always will be, a great friend.

I am extremely grateful to my parents, Bridget and Gerry and to my partner Rob for their unconditional love and support (both emotional and financial!!). To my brothers, Mark and Nigel, and my sister, Belinda, for their encouragement and interest in the progress of my research, in particular Belinda, who was continuously around to help me in any way I asked. And to my friends who were always there to listen and keep me sane!

Many thanks also to Dr. Pete Thomas who introduced me to research at the Open University and encouraged me to pursue a Ph.D. in the first place.

Last, but by no means least, thank you to my fellow colleagues and post doctoral students. To Jim and Clifford, for being generous with their extensive knowledge of the subject area and for always being available to offer informal advice. And to Rob, Sarah, Sally, Karen, Lee, Maria, Anne and Jane, who assisted in making the whole PhD process a very enjoyable experience.

DEDICATION

This thesis is dedicated to my parents, Bridget and Gerry Paine, and to Rob.

Thank you for your continuous and unconditional love and support.

CONTENTS PAGE

ABSTRACT	ii
DECLARATION	iv
ACKNOWLEDGMENTS	v
DEDICATION	vi
CONTENTS PAGE	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiv
 CHAPTER 1: Introduction	 1
1.1 Thesis Structure	3
1.1.1 Literature Review	3
1.1.2 Experimental studies	5
1.2 Terminology	6
 CHAPTER 2: Eliciting Descriptions from Children	 8
2.1 Children's Abilities as Witnesses	9
2.1.1 Developmental Considerations: Age	9
2.1.2 Cognitive Considerations: Memory	10
2.1.3 Cognitive Considerations: Language	13
2.1.4 Social Considerations: Communication	17
2.2 The Interview	19
2.2.1 The 'Typical' Interview Structure	19
2.2.2 Standard Interviewing Techniques and Children's Performance	20
2.3 Special Interview Techniques	26
2.3.1 The Cognitive Interview	26
2.3.2 Narrative Elaboration and Comprehension Monitoring	28
2.3.3 Non-Verbal Props and Prompts	29
2.3.4 Training of Interviewers	34
2.4 Chapter 2 Summary	36
 CHAPTER 3: Facial Composite Production: Developmental Issues	 38
3.1 Recalling and Recognising faces	38
3.2 The Development of Face Processing	39
3.2.1 Definitions of facial information	40
3.2.2 From Early to Contemporary Research	42
3.3 Facial Recognition	50

3.3.1 Children's Recognition Performance.....	50
3.3.2 Recognition of Faces (inner versus outer parts).....	51
3.4 Verbal Descriptions of Faces.....	52
3.4.1 Feature Saliency.....	53
3.4.2 Terms for Describing Faces.....	54
3.4.3 Verbal Overshadowing.....	58
3.5 Chapter 3 Summary.....	60
CHAPTER 4: Facial Composite Production: Practical Issues.....	62
4.1 Composite Systems.....	62
4.1.1 What is a Facial Composite?.....	62
4.1.2 Facial Composite Systems.....	63
4.1.3 The Evaluation of Composite Systems.....	66
4.2 The Composite Interview.....	71
4.2.1 The 'Typical' Composite Interview Structure.....	71
4.2.2 The Communication Process between a Witness and a Composite Operator.....	75
4.2.3 Training of the Composite Operator.....	77
4.3 Composite Systems Research.....	79
4.3.1 The Evaluation of Composites.....	79
4.3.2 Composite Systems Research with Children.....	81
4.3.3 Identification after Constructing a Composite.....	86
4.4 Chapter 4 Summary.....	88
4.5 Thesis Research Questions.....	89
CHAPTER 5: Study 1 A Questionnaire Survey of Composite Operators.....	92
5.1 Introduction.....	92
5.2 Method.....	93
5.2.1 Design.....	93
5.2.2 Participants.....	94
5.2.3 Materials.....	94
5.2.4 Procedure.....	95
5.3 Results.....	96
5.4 Discussion of study 1.....	116
5.4.1 Discussion of Study 1 Results.....	116
5.4.2 Limitations of Study 1.....	125
5.4.3 Summary: Implications of Study 1 and Future Research.....	126

CHAPTER 6: Study 2	
Children's Verbal Descriptions of Unfamiliar Facial Composites	128
6.1 Introduction.....	128
6.1.1 Research Aims of Study 2.....	131
6.2 Experiment 1: Development of Facial Stimuli for Experiment 2 (Adult Participants).....	133
6.2.1 Method	133
6.2.2 Results.....	135
6.3 Experiment 2: Children's Verbal Descriptions of Unfamiliar Facial Composites.....	137
6.3.1 Method	137
6.3.2 Results.....	143
6.4 Discussion of Study 2.....	162
6.4.1 Discussion of Study 2 Results.....	162
6.4.2 Limitations of Study 2.....	169
6.4.3 Summary: Implications of Study 2 and Future Research	171
CHAPTER 7: Study 3	
The Effect of Prompts on Children's Descriptions of Unfamiliar Faces	172
7.1 Introduction.....	172
7.1.1 Existing Prompts used by Composite Operators.....	178
7.1.2 Summary of Existing prompts and Associated Issues	181
7.1.3 Research Aims of Study 3.....	184
7.1.4 The Design and Administration of Visual and Verbal Prompts	185
7.2 Experiment 3: Adult Participants' Selection of Prompts.....	190
7.2.1 Method	190
7.2.2 Results.....	192
7.3 Experiment 4: Child Participants' Selection of Prompts	196
7.3.1 Method	196
7.3.2 Results.....	202
7.4 Discussion of Study 3.....	216
7.4.1 Discussion of Study 3 Results.....	216
7.4.2 Limitations of Study 3.....	223
7.4.3 Summary: Implications of Study 3 and Future Research	224
CHAPTER 8: Study 4	
The Effect of Prompts on Children's Facial Composites of Unfamiliar Faces	226

8.1 Introduction.....	226
8.1.1 Research Aims of Study 4.....	230
8.1.2 Composite Construction and Evaluation Considerations.....	232
8.2 Experiment 5: Composite Construction with Adult and Child Participants	233
8.2.1 Method	233
8.2.2 Results.....	241
8.2.3 Discussion of Experiment 5 Results	247
8.3 Experiment 6 – Subjective Composite Evaluation (Ranking and Rating Tasks)	253
8.3.1 Method	253
8.3.2 Results.....	256
8.3.3 Discussion of Experiment 6 Results	269
8.4 Experiment 7 - Forced Choice Matching Task.....	273
8.4.1 Method	273
8.4.2 Results.....	276
8.4.3 Discussion of Experiment 7 Results	286
8.5 General Discussion of Study 4.....	290
8.5.1 Summary of Study 4 Results.....	290
8.5.2 Limitations of Study 4.....	293
CHAPTER 9: General Discussion.....	297
9.1 Overview of Experimental Chapters.....	297
9.2 Summary of Research Findings and Theoretical Contributions.....	298
9.3 Methodological Critique.....	314
9.4 Future Directions for Research.....	321
9.5 Implications.....	326
9.5.1 Implications for Methodology	326
9.5.2 Implications for Policing Policy and Practice.....	326
9.6 Concluding Comments.....	328
REFERENCES.....	331
APPENDICES	353

LIST OF TABLES

5.1:	Factors operators reported influence their decision to ask a child to construct composite in response to Study 1	Page 99
5.2:	Mean time operators reported spending in each interview stage in response to Study 1	106
5.3:	Frequency rating of each interview component provided by operators in response to Study 1	109
5.4:	Effectiveness rating of each interview component provided by operators in response to Study 1	109
5.5:	Additional materials operators reported using in response to Study 1	111
5.6:	Reasons operators stated for showing description boxes to witnesses in response to Study 1	112
6.1	Example coding of adults' descriptors in Experiment 1	136
6.2:	Total number of matching descriptors provided by adults for each composite face in Experiment 1	136
6.3:	Example coding of children's descriptions tabulated against E-FIT description in Experiment 2	144
6.4:	Sample of summary table for categorisation of children's descriptions in Experiment 2	146
6.5:	Pattern categorisation of children's descriptions in Experiment 2	148
6.6:	Pattern categorisation for each feature element with notes for Experiment 2	149
6.7:	Mean number of descriptors provided by children in Experiment 2	151
6.8:	Mean percentage of non-verbal location and other descriptors provided by children in Experiment 2	158
6.9:	Mean number of configural descriptors as a percentage of the mean number of descriptors provided in each section of the interview by children in Experiment 2	162
7.1:	Example summary of marks awarded to children for the feature element Ear Shape in Experiment 4	194
7.2:	Example summary of marks awarded to children for the feature element Hair Length in Experiment 4	195

7.3:	The maximum score which could be awarded to child participants for each target and each set of prompts in Experiment 4	195
7.4:	The maximum inaccuracy score which could be awarded to child participants for each target and each set of prompts in Experiment 4	196
7.5:	Summary of the quantity of descriptions provided by children in Experiment 4	203
7.6:	Post-hoc Scheffé test of effect of Age on the quantity of descriptions in Experiment 4	204
7.7:	Mean number of free recall descriptions provided by children in Experiment 4	205
7.8:	Mean number of free recall person descriptions provided by children in Experiment 4	206
7.9:	Mean accuracy scores for children in Experiment 4	207
7.10:	Post-hoc Scheffé test of the effect of Age on the accuracy of descriptions in Experiment 4	208
7.11:	Mean inaccuracy scores for children in Experiment 4	208
7.12:	Post-hoc Scheffé test of the effect of Age on the inaccuracy of descriptions in Experiment 4	209
7.13:	Mean time taken (minutes:seconds) by children to select prompts in Experiment 4	210
7.14:	Children's selection of the "don't know" prompts in Experiment 4	212
7.15:	Mean accuracy scores for correct and incorrect identifications by children in Experiment 4	214
8.1:	Mean time taken (minutes:seconds) by participants to select prompts in Experiment 5	242
8.2:	Participants' selections of the "don't know" prompts in Experiment 5	244
8.3:	Percentage of participants who stated further changes to be made to the E-FIT in Experiment 5	246
8.4:	Distinctiveness ratings provided by participants for each target face used in Experiment 5	257
8.5:	Mean ranks of composites (by-item analyses) in Experiment 6	258

8.6:	Percentage of composites within each rank group (by-item analyses) in Experiment 6	259
8.7:	Post-hoc Scheffé test of the effect of Age on the ranking scores in Experiment 6	260
8.8:	Mean ranks of composite (by-participant analyses) in Experiment 6	261
8.9:	Mean ratings of composites (by-item analyses) in Experiment 6	263
8.10:	Percentage of composite receiving each rating score (by-item analyses) in Experiment 6	265
8.11:	Post hoc Scheffé test of the effect of Age on the rating scores in Experiment 6	266
8.12:	Mean ratings of composite (by-participant analyses) in Experiment 6	267
8.13:	Mean number of correct matches (by-item analyses) in Experiment 7	278
8.14:	Post hoc Scheffé test of the effect of Age on the matching scores in Experiment 7	279
8.15:	Mean number of correct matches (by-participant analyses) in Experiment 7	280
8.16:	Percentage of selections of each array photograph in Experiment 7	282
8.17:	Mean rank and rating scores for correct and incorrect identifications for Study 4	284

LIST OF FIGURES

	Page
4.1: Example of E-FIT's description 'check' box for the feature element Face Shape	73
4.2: The communication process between a witness and an operator	75
5.1: Percentage of operators who reported interviewing a child of each age in response to Study 1 (n=53)	100
5.2: Percentage of operators who stated they believe it is possible to interview a child of each age in response to Study 1 (n=43)	101
5.3: Comparison between the age of child operators reported they believe they could interview and the youngest age of child they have interviewed in response to Study 1 (n=43)	102
5.4: Comparison of the time operators reported spending with child and adult witnesses in response to Study 1	107
5.5: Operators responses to the question "Is the composite interview process too long for adults or children?" in Study 1 (n=49)	108
5.6: Percentage of operators who stated they believe the current procedure they use is optimal in response to Study 1	114
5.7: Operators responses to the question "Are composites constructed by children more, the same, or less accurate than composites constructed by adults?" in Study 1 (n=51)	115
6.1: Mean number of descriptors provided by children in Experiment 2	152
6.2: Mean number of free and prompted descriptors provided by children in Experiment 2	153
6.3: Mean numbers of descriptors provided by children for primary and secondary faces in Experiment 2	154
6.4: Mean number of descriptors provided by children for primary (left) and secondary (right) faces in Experiment 2	155
6.5: Mean number of non-verbal descriptors provided by children in Experiment 2	157
6.6: Mean number of location descriptors provided by children in Experiment 2	158
6.7: Mean number of descriptors provided by children in the comparison interview section in Experiment 2	159

6.8:	Mean number of configural descriptors provided by children in each section of the interview in Experiment 2	161
7.1:	Example of Mouth Shape prompts from the Metropolitan police	179
7.2:	Example of Average Noses from the FBI catalogue	180
7.3:	Example of Chin prompts from the Identikit catalogue	181
7.4:	Visual prompts for the E-FIT feature element of Face Shape for use in Study 3	188
7.5:	Verbal prompts for the E-FIT feature element of Face Shape for use in Study 3	189
7.6a:	Target Face 1 for Experiment's 3 and 4	191
7.6b:	Target Face 2 for Experiment's 3 and 4	191
7.7:	Mean time taken to select prompts by children in Experiment 4	211
7.8:	Percentage of correct and incorrect identifications by children in Experiment 4	213
7.9:	6-, 8-, and 10- year olds preferences for prompts in Experiment 4	215
8.1:	Mean time taken by participants to select prompts in Experiment 5	243
8.2:	Percentage of participants who provided correct and incorrect identifications in Experiment 5	245
8.3:	Mean number of changes listed by participants to be made to their E-FIT in Experiment 5	247
8.4:	Facial images of target 5 for ranking and rating tasks	254
8.5:	E-FIT of Target 2 correctly matched by all participants in Experiment 7	277
8.6:	E-FIT of Target 1 frequently incorrectly matched to Target 4 in Experiment 7	283
8.7:	E-FIT of Target 3 frequently incorrectly matched to Target 4 in Experiment 7	283

CHAPTER 1: INTRODUCTION

Obtaining eyewitness evidence from children about events they have experienced is a problematic process. Typically, psychological research has focused on the dilemmas inherent in interviewing children who are suspected victims of child abuse (e.g. Poole and Lamb, 1998), where identification of the alleged perpetrator is rarely an issue. However, although stranger-perpetrated child crimes are less common than non-stranger abuse, they occur with sufficient frequency to warrant concern over children's abilities: "a person's ability to describe or recognise faces accurately can be a critical forensic factor when the person has witnessed a crime or been the victim of actions by a stranger" (Brigham, 2002, p.115). When the identity of an offender is unknown in a criminal investigation witnesses are often asked to produce an external representation of the suspect's face from their memory with a police operator. This external representation can be a verbal description and/or a visual representation (or facial 'composite' image), which can then be used for circulation to other police forces or to the general public.

From an empirical perspective, the process of creating an external representation utilises all facets of the face processing system and memory: witnesses are required to describe, recall and recognise the face of the perpetrator during the process of facial reconstruction, therefore making composite construction an ideal method for studying face processing in humans (Pike, Kemp and Brace, 2000). Brigham (2002) commented: "knowledge about how children perceive faces and how accurately they identify and describe them, and an awareness of factors which may affect this accuracy can be

important for evaluating the facial descriptions and identifications children may give during interviews” (p.115).

Given the importance of verbal descriptions and facial composite images in the criminal justice system, it is interesting to note that in contrast to the wealth of research on interviewing child witnesses, the application of this research to obtaining verbal descriptions and facial composites of unfamiliar faces from children has been somewhat neglected, both by legal systems and by psychological research. Schwartz-Kenney, Bottoms and Goodman (1996) stated that little is known about children’s abilities to report pertinent information about the appearance of unfamiliar people, and even less is known about how to obtain accurate descriptive reports from children. Flin, Davies and Stephenson (1987) suggested that although this kind of information is collected from child witnesses “we know very little about children’s abilities to perform such tasks. Despite the fact the descriptive statements and positive identifications can constitute the corner stone of a police investigation and subsequent prosecution” (Flin *et al.*, 1987, p.281). Research which has been conducted has shown that neither adults’ or children’s competence at composing an external representation is high (Davies, 1996). However, despite this inability of witnesses (both adults and children) to produce accurate composites we must not necessarily conclude that they do not have the underlying skills necessary for such a task - perhaps it is just the way the task is presented that causes difficulty (Sporer, 1996).

In some cases, children have shown particular prowess at producing verbal descriptions and visual representations of unfamiliar faces (Davies, Shepherd, Shepherd, Flin and Ellis, 1986). An example of the abduction of a 6-year-old girl in Staffordshire illustrates such prowess. The 6-year-old child drew a picture of her abductor, which featured on

the BBC television programme *Crimewatch UK*, along with a facial composite (constructed using the E-FIT composite system¹) of the abductor (produced by a local man). Her drawing shows the abductor wearing a woolly hat, with a tuft of hair over the forehead. The Detective Inspector in charge of the case stated that the child was able to provide some “significant details” (see ‘Girl draws picture of abductor’, 2001).

The need exists then for research which broadens the understanding of children’s verbal descriptions and visual representations of unfamiliar faces and gains a better understanding of the facial processing skills of children and the effects they may have on composite construction. Therefore, the central aim of this thesis was to examine and apply developmental, eyewitness, facial recognition, facial recall and composite research with children to determine how they might be assisted, through appropriate interview techniques, to provide facial descriptions and composites of an unfamiliar face.

1.1 THESIS STRUCTURE

1.1.1 LITERATURE REVIEW

The review was structured to address each of the main concerns highlighted in the literature: from children’s general abilities as witnesses to their ability to provide facial descriptions and to produce facial composites. Additionally, the effects of interview techniques (in particular ‘special’ interview techniques) on these abilities were considered. The importance of an interview as a two way communication process between a child witness and an interviewer is discussed throughout the thesis.

¹ The E-FIT composite system is described in Section 4.1.2.

- **Chapter 2: Eliciting Descriptions from Children**

The increased involvement of children in the criminal justice system requires a thorough understanding of children's abilities as witnesses and has renewed interest in how one should interview child witnesses in order to elicit descriptions (e.g. Aldridge and Wood, 1998; Wilson and Powell, 2001). Chapter 2 introduces children's abilities as witnesses from a general perspective. It covers both the importance of interview techniques and provides a developmental backdrop against which children's abilities as witnesses should be judged.

- **Chapter 3: Facial Composite Production: Developmental Issues**

As mentioned above, the process of creating a facial composite accesses all facets of the face perception system. Chapter 3 considers the developmental issues surrounding facial descriptions, recall and recognition.

- **Chapter 4: Facial Composite Production: Practical Issues**

After considering children's general abilities as witnesses (Chapter 2) and the developmental issues of facial composite production (Chapter 3), Chapter 4 considers the more practical issues surrounding the procedure of obtaining facial composites from child witnesses.

Each of the above chapters ends with a summary of the main issues. At the end of Chapter 4 the main research questions that formed the impetus for the current research are listed.

1.1.2 EXPERIMENTAL STUDIES

Analogous with the literature review, the experimental studies were structured to address children's abilities to provide facial descriptions and to produce facial composites. A particular emphasis of the studies was on the creation and evaluation of a developmentally appropriate interview technique for facial interviews with children. Detailed methodological information and the principle results of each experiment are presented in Appendix I.

- **Chapter 5: Study 1 - A Questionnaire Survey of Composite Operators**

Study 1 involved a questionnaire survey of facial composite police officers (operators) in the United Kingdom (UK) to identify their current practice with, and experiences and opinions of, child witnesses (Paine, Brace, Pike, and Westcott, 2002, 2003b).

- **Chapter 6: Study 2 - Children's Verbal Descriptions of Unfamiliar Facial Composites**

Study 2 explored the content of children's verbal descriptions of unfamiliar faces in order to establish the language and terms children use (Paine, Brace, Pike, and Westcott, 2003a). The children's descriptions obtained in this study were utilised to produce an original set of age appropriate verbal and visual prompts, which were then used to interview children in Studies 3 and 4.

- **Chapter 7: Study 3 - The Effect of Prompts on Children's Descriptions of Unfamiliar Faces**

Study 3 investigated the effect of the verbal and visual prompts as potentially appropriate interview techniques on the production of children's descriptions of unfamiliar faces (Paine, Brace, Pike, and Westcott, 2004).

- **Chapter 8: Study 4 - The Effect of Prompts on Children's Composites of Unfamiliar Faces**

Study 4 examined the effect of the verbal and visual prompts as potentially appropriate interview techniques to assist child participants in their composite constructions of unfamiliar faces (Paine *et al.*, 2004).

The final chapter (**Chapter 9**) is a general discussion of the thesis.

1.2 TERMINOLOGY

A final note on terminology is required:

- **Child Witness**

The definition for the term 'child witness' is based on definitions provided in the British Government's guidance *Achieving Best Evidence in Criminal Proceedings: Guidance for Vulnerable or Intimidated Witnesses, including Children* (Home Office, 2002) (hereafter referred to as ABE). The unqualified term 'child' or 'children' refers generally to children of all ages up to the upper age limit defined in the 1999 Youth Justice and Criminal Evidence Act (i.e. below 17-years of age). If further clarification is required, the term 'very young children', refers to children of nursery school age (i.e. up to 5-years of age). The term 'young children' refers to children of primary school age

(i.e. up to 11-years of age), while ‘older children’ denotes those of secondary school age (i.e. over 11-years of age). The experimental chapters of this thesis (Chapters 5 to 8) focus on ‘young children’ (i.e. children aged from 6- to 10-years).

- **E-FIT**

Electronic Facial Identification Technique (E-FIT) is a computerised facial composite system. It is currently the most commonly used computer based system for constructing composites in use by police forces in the UK (Clark, 2000) and therefore was used as the facial composite system in the experimental chapters of this thesis (Version 3.1a for Windows).

- **Facial Composite**

The term ‘facial composite’ is defined as an image made up of other images or parts. It may be computer generated (by a police composite operator) or hand drawn (by a police sketch artist) (Clark, 2000).

- **Witness**

The term ‘witness’ is used to include both witnesses and victims.

CHAPTER 2: ELICITING DESCRIPTIONS FROM CHILDREN

Before the specific theoretical and practical issues of facial descriptions and composite systems are addressed (Chapters 3 and 4) it is necessary to consider children's abilities as witnesses from a general perspective. This chapter provides a developmental backdrop against which children's abilities as witnesses should be judged and covers the importance of interview techniques (the structure of the interview, questioning techniques, including special questioning techniques). The communication process between the child witness and the interviewer will be focused on throughout the chapter.

In a recent high profile and politically sensitive investigation, four boys, aged between 15- and 17-years, were cleared of murdering 10-year-old Damilola Taylor (see 'Damilola Taylor Verdict', 2002). The police and prosecution in charge of the investigation focused on a single witness: a 14-year-old girl. The policewoman who questioned the witness was criticised for her unprofessional treatment of the witness. During the trial the judge criticised the police for not following Home Office guidelines, set out in the British Government's *Memorandum of Good Practice on Video Recorded Interviews with Child Witnesses for Criminal Proceedings* (hereafter referred to as MoGP) (Home Office, 1992), on how a potential child witness should be interviewed. The 14-year-old girl was also suggested by the press to be dishonest and unreliable (The Times, April 26, 2002). This case highlights the confusion concerning both children's capabilities as witnesses and interviewers' abilities to question children.

2.1 CHILDREN'S ABILITIES AS WITNESSES

A number of detailed reviews and commentaries into children's abilities as witnesses have been published (e.g. Ceci and Bruck, 1993, 1995; Poole and Lamb, 1998; Westcott, Davies and Bull, 2002). These reviews emphasise the importance of findings from developmental, cognitive and social psychological research for the improvement of child witness interviewing. Furthermore, ABE warns police and court staff to take into account the level of "cognitive, social and emotional development" when conducting interviews (Home Office, 2002, p.2-3).

2.1.1 DEVELOPMENTAL CONSIDERATIONS: AGE

Age has long been identified as a general indicator of an eyewitness's ability. "In the past, children were often ignored by legal systems because of the belief that anyone below the age of around 10-years was 'incompetent'" (Bull, 2001, p. xi). When the performance of very young children is compared to that of older children and adults, it is usual to find age differences in both the completeness and the accuracy of reports (Ceci and Bruck, 1993). A heuristic identified in the literature is that the age 7-years represents a significant shift in children's abilities and associated developmental changes (Westcott *et al.*, 2002).

However, there are consistent findings from the research which posit that age alone does not determine ability to provide complete and accurate accounts of events. "Age is only one factor in the complex determination of performance and other variables can be equally if not more significant" (Davies, 1996, p. 253). Rather than young children having difficulty in the recollection of information, it may be that the interaction between age and other, environmental factors affects the testimony of young children disproportionately. For example, Jones (2003) stated that children under 10- to 11-years

are more susceptible to social pressures, both because of a wish to please and comply with the perceived demands of adults around them, and because of their reaction to adult authority. Developmental differences are considered throughout this thesis.

2.1.2 COGNITIVE CONSIDERATIONS: MEMORY

Given its linkage with testimony, a primary cognitive consideration in children's abilities as witnesses is memory. Memory is not a single entity but is a collection of systems and processes. Bornstein and Lamb (1992) defined memory as the process of storing what is attended to, and then being able to retrieve and use that information. A brief overview of the three stages of memory that can be distinguished are described below:

Stage 1: The Acquisition Stage

This stage involves the perception and registration of an event in which information is encoded. Reviews of the literature (e.g. Gathercole, 1998) suggest that with age children become increasingly efficient at developing ways to encode, store, rehearse and subsequently retrieve information. In terms of *encoding*, children and adults attend to different environmental features and hence encode different information into memory. In general, young children attend to a greater variety of stimuli in their environments and are easily distracted by 'irrelevant' stimuli (Lepore, 1991). Identifying information, such as Eye Colour² or Height may be noticed and recalled by an adult witness but overlooked by a child witness, who instead recalls something more salient to them, but forensically less relevant (King and Yuille, 1987). However, occasionally younger children's scattered attention may enable them to encode information that is relevant to an interrogator and missed by older children and adults (Lepore, 1991).

² Due to the importance of facial (and person) features in this thesis these are referred to with capital letters.

Stage 2: The Storage Stage

This stage entails the committing of information to either short term or long term storage. In terms of storage, relative to older children and adults, young children have less capacity and a more constrained knowledge base for efficiently organising information stored in memory (Lepore, 1991). Memory is affected by the circumstances at the time of an event occurring (Jones, 2003) and an individual child's ability to encode and store information accurately may also be considerably affected by stress. The impact of stress on children's memory has been the subject of much debate amid contrasting findings (e.g. Goodman, Bottoms, Schwartz-Kenney and Rudy, 1991; Peters, 1991).

Memory is also affected by the passage of time. The amount of forgetting likely to take place over different delay intervals is not yet well known (Sporer, 1996) and few studies have examined whether children's memories are more sensitive to longer delay intervals than when compared to adults. In a study conducted by Flin, Boon, Knox and Bull (1992) children aged 6- and 9-years of age, and adults, were interviewed one day after witnessing a staged event and then again five months later. Results showed that while witnesses of all ages forgot information over this period, children recalled slightly less information than adults, and also showed a greater reduction in recall accuracy (particularly the youngest age group of children). Similar results were reported by Poole and White (1993) in a study of 4- to 8-year-olds and adults over a longer period. Children and adults reported comparable proportions of information one week after

witnessing an event. However, two years later, the proportion of inaccurate information reported by children increased whereas adults maintained their earlier error rate³.

Stage 3: The Retrieval Stage

This final stage concerns the act of retrieving stored information and is the most important stage in relation to this thesis. Even if a child has encoded an event, young children have difficulty spontaneously retrieving the information from memory (Salmon, 2001) and are less flexible in generating and using strategies to search their memory efficiently (Flavell, Miller and Miller, 1993). It is not until a child reaches the age of 8-years that they begin to employ the more complex and successful independent retrieval strategies of older children and adults (and even then not to their fullest effect). Hence younger children may require more prompting, in the form of external retrieval cues, than older children do.

In summary, studies have not found a simple relationship between memory and age. However, a general heuristic in the literature asserts that by age 8-years, children's capacity to encode, store and retrieve information is on a par with that of adults (Jones, 2003).

Davies (1996) stated that although important to memory development, effective coding and retrieval strategies are only a partial guide to the likely competence of child witnesses. Children may remember incidents but a variety of factors influence the quality of the information they provide. These factors are discussed in the following

³ At the time of writing there is no official data concerning the delays experienced by child witnesses waiting to be interviewed before they give evidence in court or provide facial descriptions and/or composites. Research studies conducted in Scotland have reported that child witnesses wait an average of seven months before giving evidence in court (Flin, Bull, Boon and Knox, 1990). However, not all children who provide descriptions of strangers may do so after such long delays. The 'typical' delay experienced by the Metropolitan police has been estimated at around one week (A. Parry, personal communication, 14Th November, 2001).

sections along with additional cognitive factors that have been shown to develop with age.

2.1.3 COGNITIVE CONSIDERATIONS: LANGUAGE

The legal system expects that witnesses possess sufficient language skills to translate their memories into verbal responses to convey what happened in a question-answer format (Saywitz and Snyder, 1993). However, there is a difference between the ability to recall and the ability to report what one recalls. Language development places limits on the extent to which children can communicate their experiences (Harley, 1996) even if the information is available, and the typical language of legal professionals questioning witnesses is often developmentally mismatched with the abilities of the children answering (Perry, McAuliff, Tam, Claycomb, Dostal and Flanagan, 2001). For the interviewer, it is important to evaluate both children's *production* and *comprehension* of language.

Children's Production of Language

Knowledge about language production is important in order to accurately interpret young children's speech. Research has shown that even by the age of 5- or 6-years, children do not verbally recode aspects of their experiences, resulting in the possibility that information may be stored yet not available for verbal reporting (e.g. Price and Goodman, 1990). Children's vocabulary expands faster than their comprehension of language (Walker, 1999), therefore children frequently use words before they know their conventional adult meaning. Although a common vocabulary is assumed between adults, this cannot be extended to children as they may have their own terms or may use a word differently from an adult (Saywitz, 2002). Additionally, children can include a broader range of meanings within their use of one word, or alternatively a much more restricted use (Jones, 2003). It is common for interviewers to interpret children's terms

incorrectly and, as children rarely ask for clarification, they are more likely to simply accept an interviewer's interpretation (Saywitz, 2002). Ideally, interviewers should avoid introducing new words until a child has used the word first (Home Office, 2002) and should "try to pick up on the 'labels' children use for various incidents and use these in questioning" (Home Office, 1992, p. 19).

Children who have experienced events outside their usual range of experience may not have the vocabulary to describe what has occurred; for example, younger children typically do not have the vocabulary to describe sexual experiences (for a review see Volbert and van der Zanden, 1996). Legal events are also unfamiliar to children. Saywitz, Jaenicke and Camparo (1990) suggested that normative data should be gathered on age-related patterns of the use and understanding of legal terms commonly encountered by child witnesses. In the specific area of facial recall, the majority of research has concentrated on collecting adults' terms for describing faces at Aberdeen University in the 1970s and 1980s (e.g. Shepherd and Deregowski, 1981). Research has only just begun into children's vocabulary for faces and there have been no published studies to date on children's understanding of terms. This is discussed further in Chapters 3 and 4.

Children's Comprehension of Language

In addition to language production, children's language comprehension develops rapidly with age (Walker, 1999). When adults question children there may be an assumption that children will not answer a question if they do not understand it. However, children can differ in their ability to comprehend linguistically complex constructions and if young children do not have a full syntactic or semantic understanding of the linguistics of a question then their recall will be affected (Saywitz, 2002). Poorly worded questions "may be a potentially serious source of

miscommunication” (Walker, 1993, p.59). However, if the ‘right’ questions are asked in the ‘right’ way, Walker (1999) believes that even very young children can tell what they know.

Studies of language development and acquisition (see a review by Foster-Cohen, 1999) suggest that there are a limited number of words children can process in a sentence and that lengthy compound sentences with embedded clauses and other linguistic complexities may be beyond the comprehension of children under 8-years of age (Saywitz, Nathanson and Snyder, 1993). Such utterances should be broken down into several short questions in order to elicit reliable information from a younger child. When confronted with such linguistically complex questions and sophisticated vocabulary that exceed children’s level of language competence, young children have limited and ineffective methods of coping with instances of non-comprehension (Saywitz, Nathanson and Snyder, 1993). Strategies for coping with non-comprehension develops gradually with age although even school age children rarely ask for clarification or indicate misunderstanding, instead, they often try to answer questions they do not fully understand (Saywitz and Snyder, 1993). Studies have shown that children are willing to answer nonsensical questions without any understanding of what they are being asked, and their answer cannot be taken as evidence that they understand the question (e.g. Hughes and Grieve, 1980). Waterman, Blades and Spencer (2000) asked 5- to 8-year-olds sensible and nonsensical, open and closed questions. Closed nonsensical questions include the example “is red heavier than yellow?”, open nonsensical questions include the example “one day there were two people standing at a bus-stop. When the bus came along who got on first?”. Waterman *et al.* (2000) showed that, in contrast to previous research, children do not answer all nonsensical questions. They found that the way a question was phrased had a significant effect on whether or

not children indicated when they did now know the answer. Children answered all the sensible questions appropriately, and only attempted to answer a small proportion of the nonsensical open questions. However, they did try to answer three-quarters of the nonsensical closed questions and this is important given the findings that when children are told that they may ask for information to be clarified they have been shown to rarely do so (Markman, 1977, 1979).

Children also differ in their ability to communicate their understanding of commonly used legal terms. Saywitz *et al.* (1990) used transcripts from legal proceedings to produce a list of legal terms which were used frequently in cases involving child witnesses. Results showed that many common legal terms were unfamiliar to or misinterpreted by children under 10-years of age. Saywitz *et al.* (1990) concluded that “age appropriate word choice in the examination of child witnesses may be an important factor in eliciting accurate testimony” (p.531).

Children's Use of “I don't know” as a Response

As mentioned previously, children tend to try to answer a question rather than saying “I don't know”. A number of studies have been conducted in which children from 6- to 11-years of age have been either instructed or trained to respond with “I don't know” when uncertain of answers to questions posed by adults. The aim of the studies was to increase the accuracy of responses. Saywitz, Nathanson, Snyder and Lamphear (1993) trained children aged 6- to 11-years of age to say “I don't know” when responding to an eyewitness task. Training involved children listening to a story about a child who agreed with the suggestions implied by the questions of adults and the negative effects of this. Results showed that training reduced the proportion of incorrect responses to misleading questions, even for 6-year-olds. Nesbitt and Markham (1999) examined a similar training with 4-year-olds and clearly demonstrated that trained children were

more accurate in their responses than untrained children. Finally, Mulder and Vrij (1996) found that explaining that “I don’t know” is an acceptable answer; and that an interviewer could not help answer questions reduced suggestibility in 4- to 10-year-olds.

All of the studies cited above support the conclusion that before young children are questioned, it would seem advisable to give them a short training program in the importance of saying “I don’t know” when they are uncertain of an answer. Furthermore, the British Government’s current guidelines for interviewing child witnesses (ABE) encourages interviewers to reassure a child they can say when they do not understand a question or if they do not know the answer to a question (Home Office, 2002).

2.1.4 SOCIAL CONSIDERATIONS: COMMUNICATION

The limitations of children’s abilities to both *comprehend* and *produce* language may lead to a communication breakdown between a child witness and an adult interviewer⁴. Language and communication are about social relationships as well as the communication of understanding and memory and communication is very much a two way process. Mercer (2000) states that language “provides us with a means for thinking together, for jointly creating knowledge and understanding” (p.15).

Due to the way in which children have been socialised to communicate with adults, they rarely volunteer detailed information. In a forensic interview, a child is in an unusual situation, they must understand that the adult is asking a question because s/he does not know the answer to it (Wattam, 1992). A child may not believe that they are providing information which will aid a criminal investigation (Nesbitt and Markham, 1999).

⁴ The communication process between a witness and an interviewer is described in more detail in Chapter 4.

Factors present during interviews can predispose a child to respond to a question in a particular way, for example, children may also respond with an answer that is expected to please the adult regardless of whether the child knows the answer to be correct or not (Moston, 1990). Until communicative competence is fully developed, miscommunication can be an obstruction to eliciting reliable information from children (Saywitz, Nathanson and Snyder, 1993). Saywitz (2002) stated that adult-like communicative competence is not fully developed until age 10- to 12-years.

An overview of this kind helps to generate expectations of children's abilities as witnesses. However, it can obscure the considerable individual differences between children. Although consistent findings of age differences across studies exist, it is important to consider the vast differences between individual children (Bruck and Ceci, 1999).

One of the most important shifts in thinking about children as witnesses has been the movement from examining whether or not children are competent witnesses, to determining what conditions children are most, or least, competent under (e.g. Westcott *et al.*, 2002). Therefore, the responsibility has shifted from being placed solely on the child to a sharing of the responsibility with the adult interviewer for the provision of accurate information. Saywitz's (1995) comment that interviewers must "capitalise on their (children's) strengths, compensate for their weaknesses, and create an optimal environment for their remembering and communicating" (p. 113) reflects this. Research which has focused on the impact of interviewers' style of questioning (Ceci and Bruck, 1993; Poole and Lindsay, 1995), and the use of special interview techniques on children's descriptions will be considered in the following sections of this chapter.

2.2 THE INTERVIEW

2.2.1 THE 'TYPICAL' INTERVIEW STRUCTURE

Although every interview is highly individual, there is substantial consensus about how forensic interviews should be conducted⁵. A number of guidelines have been introduced to provide guidance for interviewing child witnesses. These guidelines include the British Government's MoGP (Home Office, 1992) which has since been replaced by ABE (Home Office, 2002). The majority of interview guidelines published see interviews progressing through a series of stages. A summary is described below of the main stages of an interview within the three main published interview guidelines for use with child witnesses and victims (ABE, Home Office, 2002; The Cognitive Interview, Fisher and McCauley, 1995; and the National Institute for Child Health and Development (NICHD) structured protocol for investigative interviews, Lamb, Orbach, Sternberg, Esplin and Hershkowitz, 2002):

Stage 1 - Establish Rapport

It is generally accepted that the first stage of any interview should be a preliminary discussion between a child witness and the interviewer in order to provide an introduction, establish rapport and gauge a child's social, emotional and cognitive development (e.g. Boggs and Eyeberg, 1990). During this stage, the purpose of the interview is explained and an interviewer will clarify the ground rules (describe the child's task and explain that the child can say when they do not understand and that they may correct the interviewer). A child may also participate in a practice interview (Poole and Lamb, 1998) in order to establish the ground rules and to allow children to practice indicating that they do not understand or do not know the answer to a question (e.g. Mulder and Vrij, 1996).

⁵ The 'usual' structure for a facial composite interview is described in Chapter 4.

Stage 2 - Free Descriptions

The second stage of an interview requires a witness to provide a free description of the relevant event(s) in his or her own words. Free descriptions are “a measure of retention in which no external help is given to the recaller in his attempt to retell” (Clifford and Bull, 1978, p.19). Interviewers are urged to listen carefully to a child’s responses and whenever appropriate to use facilitators (Lamb *et al.*, 2002) which are neutral encouraging statements such as ‘uhuh’, ‘o.k.’, or ‘yes’ and to repeat verbatim the last words of the child’s proceeding response to get a child to report as much as possible before the next stage of asking questions.

Stage 3 - Questioning

Questioning may take on a number of different forms. Recommendations state that *open* questions should be used initially, followed by *directive* (specific or focused) questions. *Option posing* questions (closed or force choice) questions should be used as seldom and as late as possible in the interview, and *suggestive* questions (leading or misleading) should be avoided. Again, interviewers should use general prompts and minimal encouragement. Directive questioning may be aided by using special interview techniques⁶.

Stage 4 - Closure

As the interview draws to a close the interviewer should summarise the evidentially important statements made by the child and answer any questions from the child. The interview should return to neutral topics to leave the child in a positive state of mind.

2.2.2 STANDARD INTERVIEWING TECHNIQUES AND CHILDREN'S PERFORMANCE

Due to the importance of obtaining complete and accurate information from children, a considerable number of studies (in both laboratory and field settings) have extensively

⁶ ‘Special’ interview techniques are described in section 2.3.

explored the association between the standard interview techniques and the quantity and quality of information provided by children (e.g. Goodman and Reed, 1986; Lepore, 1991). Such studies have identified a continuum of risk for questioning: ranging from open questions at one end, through directive questions, option posing questions to suggestive questions at the other end of the continuum. Evidence from the studies which have been conducted to date and children's related performance for each of these question types (based on Lamb *et al.*, 2002) are now considered.

Open Questions and Free Descriptions

Open questions or statements require open-ended responses from a child (e.g. "tell me what you can remember about the man you saw" and/or "and then what happened?"). These open questions are neither contaminating nor suggestive and it is well recognised that the open responses or free descriptions/recall they elicit represent the most accurate form of recall in children and adults in both laboratory and forensic contexts (Dent, 1986; Orbach and Lamb, 2000). Goodman and Reed (1986) found children of 3-years and 6-years of age to be no less accurate than adults in their free descriptions. Leippe, Romanczyk and Manion (1991) confirmed the same trend.

The encouraging findings on the quality of children's recall needs then to be balanced by the differences in quantity of information provided. Legal and psychological evidence indicates that although young children's free recall of an event can be quite accurate, it tends to contain little detail (Lampien and Smith 1995). Schwartz-Kenney, Bottoms and Goodman (1996) stated that in response to free recall questions about the appearance of a person, children provide little information. This may mean that, in many cases, if investigators have no other relevant information about a suspect's identity, children may not be able to provide enough initial information to guide authorities. There is a consensus among researchers that the quantity of children's free

recall increases with age (for a review see Davies and Westcott, 1999). For example, in Leippe *et al.*'s (1991) study adults' spontaneous statements were up to three times as long as the statements from 5 to 6-year-olds. Children's ability to provide only small amounts of information in response to a free recall task is due in part to their limited ability to use retrieval strategies and the demand for fairly advanced language production abilities that are not well developed in young children as described previously. Therefore, such a task may underestimate children's actual memory of an event. Lepore (1991) reported that the amount of details spontaneously recalled by children increased up to the age of 11-or 12-years and then approached adults' level of recall.

In order to build on the narrow base of information provided by free descriptions, it is necessary for an interviewer to question a child in order to elicit the fullest possible accounts (Orbach and Lamb, 2000). With appropriate scaffolding (Bruner, 1976), or support from adults, children from the age of 2 have been shown to have a verbal, albeit limited, recall of events (Westcott *et al.*, 2002). Studies show that the completeness of free recall reports increase with the level of prompting used (e.g. Dent, 1992).

Many interviewers use more recognition memory prompts and even leading (suggestive) prompts when questioning young children, even though such prompts are more likely to elicit inaccurate information. There are concerns about children making an increasing number of errors in prompted, or cued, recall situations. However, the less information that is initially provided by a witness, the greater the number of questions that may be asked and, as Lepore (1991) stated, children will tend to make more errors in their cued recall than adults because they are routinely asked more questions to compensate for their inadequate free recall.

Directive Questions

Ideally, only after open questions are no longer productive do interviewers proceed to directive (specific/focused) questions addressing details previously mentioned by the child. Directive questions should begin by concentrating on one or more specific aspects of the witness's free description (Poole and Lamb, 1998). These can be cued invitations (e.g. "tell me about that knife") or 'wh-' questions (e.g. "what colour was his hat?", after the child mentioned a hat).

Directive questioning may encourage the child to provide definite answers to questions about which s/he is unsure and produce more information than open questions because they can serve as cues for the witness, which can trigger a child's memory about an event (Steller and Boychuk, 1992). Lepore (1991) reported that performance on such tasks approaches adult levels by the age of 6- or 8-years. However, depending on the type of directive question, the increase in the quantity of information provided can be at the expense of the quality of the information recalled (e.g. Dent and Stephenson, 1979). Although responses to directive questions are not as accurate as to open questions, they are more accurate than responses to suggestive questions (described below) (Lamb *et al.*, 2002).

Option Posing Questions

If directive questioning proves unproductive, interviewers may move onto option posing (closed/forced choice) questions⁷. These questions focus a child's attention on aspects of an event a child had not previously mentioned (e.g. "did you see a knife?"). Guidelines recommend that option posing questions should never be used for probing central events in the child's account which are likely to be disputed at court, and should be reserved until the very end of the interview (Bull, 1995).

⁷ Option posing questions were called 'leading' in some previous reports but were relabelled by Lamb *et al.* (2002) to avoid confusion with questions described as leading by other professionals.

Option posing questions can either require a yes or no response from the witness, (e.g. “was the man wearing a hat?”) or they can be multiple choice (e.g. “was the hat blue or green?”). These questions seem to be interpreted by young children as calling for a response, even when children have no idea what the question is asking. Children may guess which response is expected of them. In part this may be because children attempt to be co-operative conversational partners (Ceci and Bruck, 1995). As described in Section 2.1.4, a child may believe that the adult knows the answer and may therefore respond with an answer that is expected to please the adult interviewer.

Overall, research has shown that children answer forced choice questions less accurately, and are more suggestible to them (e.g. Poole and White, 1993). An analysis of transcripts of American investigative interviews by Hunt, Komori, Kellen, Gallas and Gleason (1995) showed that 74% of children responded simply with one choice or another to forced choice questions rather than providing their own answer. Specifically, very young children are more likely than older children to respond erroneously to forced choice questions (Ceci and Bruck, 1995; Poole and Lindsay, 1998). Waterman, Blades and Spencer (2000) showed that if 5- to 8-year-old children were asked nonsensical or unanswerable questions which only required a yes or no response, then children tended to give inappropriate responses. However, if the question was phrased in an open format, the majority of children indicated they did not understand or did not know the answer.

On a more positive note, some research has revealed that use of forced choice questions does not have a detrimental effect on the quality of facial descriptions obtained from children (Schwartz-Kenney, Bottoms and Goodman, 1996) or on the likeness of

composites constructed with adults (Clark, Pike, Brace and Kemp, 2001), as long as a “don’t know” option is included. These studies will be described in more detail in Section 7.1

Suggestive Questions

Suggestive (leading/misleading) questions are stated in such a way that the interviewer strongly communicates what response is expected or assumes details that have not been revealed by the child. Research has shown that children may agree with the information suggested to them and may incorporate it into their descriptions (Goodman, Rudy, Bottoms and Aman, 1991). The adverse effects of suggestive (and option posing) questions are exaggerated when they occur early in the interview and when the children being questioned are very young (Saywitz and Goodman, 1996). If young children’s memory is weak, it is particularly susceptible to the effects of suggestion during retrieval. Research proposes that the effects of suggestion decrease with age: by the age of 8- or 9-years children are less likely to comply with suggestive questioning, and children older than this can be as resistant as adults.

Repeated Questioning

Legal testimony often involves repetitious questioning both within interviews (to elicit more detailed responses) and between interviews (following long delays or when a witness is interviewed by different professionals). It is assumed that answers to repeated questions will be less accurate than original answers because witnesses react to the social pressures of repeated questions by offering speculative responses (Moston, 1992). An important distinction to make is that repeated questioning can be both open (asking for more information without suggesting a particular answer) or closed (suggesting a particular answer). Both types of repeated questions may have different effects on children of different ages. Research supports the use of repeated questioning as a probe for more information to open ended responses. Poole and White (1991) examined 4- to

6-year-old children's answers to repeated sets of open and closed questions about a novel event. They found children were as accurate as adults with open ended questions, however, 4-year-olds were more likely to change responses to repeated closed questions. Memon and Vartoukian's (1996) findings with 5- to 7-year olds showed a similar pattern. Perhaps the lack of an age effect with open questions is due to the fact that children are used to open ended questions being repeated in daily interactions (e.g. "what did you say you did at school today?") and children respond to these sorts of questions by repeating their original story. Age effects have also been reduced following a long delay between an event and subsequent interview. For example, Powell and Thomson (1997) examined 4- to 5-year-olds and 6- to 8-year olds memory of the final occurrence of a repeated event and found age differences were less evident at a six week than at a one week retention interval.

In summary, research has shown children's performance varies depending on the form of questioning employed by interviewers. As described at the end of Section 2.1.4, developments in interviewing techniques have been heavily influenced by the findings of the psychological research described in order to optimise children's recall. Some of these 'special' interview techniques are now considered⁸.

2.3 SPECIAL INTERVIEW TECHNIQUES

2.3.1 THE COGNITIVE INTERVIEW

The Cognitive Interview (CI) was originally developed by psychologists Geiselman and Fisher (1992) to improve police interviewing procedures by efficiently enhancing eyewitness recall. The CI evolved in two distinct phases from 1985. The original CI was

⁸ A composite interview could also be viewed as a 'special interview technique' and will be discussed in more detail in Chapter 4.

based on a limited set of interview principles to increase memory retrieval in ideal conditions based on laboratory studies. The enhanced CI was developed by Fisher, Geiselman, Raymond, Jukevich and Warhaftig (1987) and has been integrated into UK police interviewing guidelines as a powerful interviewing technique (Clifford and George, 1996). It consists of a package of mnemonic techniques designed to assist witnesses in searching their memory more exhaustively through multiple attempts at recall within a single session (Memon, 1998). The CI primarily focused on the improvement of recall of details of actions and events but has been applied to improve the descriptive details of suspects.

The CI was originally developed for use with adult witnesses and has repeatedly been shown to increase levels of recall relative to traditional interviewing, without a rise in error (e.g. Fisher and McCauley, 1995; Koehnken, Milne, Memon and Bull, 1999). The CI has also been adapted for use with children, however, the results are not as clear as with adults.

Fourteen studies (published and unpublished) have examined the use of the CI with children from the age of 6-years onwards (see review by Milne and Bull, 1999) and most of these have demonstrated positive effects (e.g. Geiselman and Padilla, 1988; Geiselman, Saywitz and Bornstein, 1990). Only two studies have found no effect of the CI on the quantity of information recalled (Memon, Cronin, Eaves, Bull, and Kupper, 1993; Memon, Wark, Holley, Bull and Koehnken, 1996). Although there are a number of criticisms regarding the use of the CI with children and various researchers suggest that some of its elements should be used with great caution, for example, Saywitz, Geiselman and Bornstein (1992) speculated that the 'change-perspective' techniques should be omitted from the interview with young children. Other aspects of the CI do

not appear to cause any problems, even for very young children (e.g. 'reinstate context' and 'report everything').

2.3.2 NARRATIVE ELABORATION AND COMPREHENSION MONITORING

The Narrative Elaboration (NE) Procedure (Saywitz and Snyder, 1993, 1996; Saywitz, 1995) was designed to help children overcome the developmental limitations in memory language and communication described in Sections 2.1.3 and 2.1.4. NE focuses on pre-interview training of witnesses, where interviewers allow children to practice narrative elaboration techniques and encourage them to correct interviewers before they begin structured questioning. An interviewer trains the child in the use of 'reminder' cue cards which contain a generic pictorial representation of a story-grammar category representation of four story-grammar categories: (1) participants; (2) setting; (3) actions; and (4) conversations and affective states relevant to a target event. The interviewer then begins questioning with a free-recall question, followed by presentation of each cue card accompanied by the open question, "Does this card remind you to tell anything more?", followed by a list of directive questions.

Researchers have examined the effectiveness of the NE procedures with young children, and results showed that during cue-card presentation, children interviewed with the NE procedure reported a greater amount of accurate information, but not a greater amount of inaccurate information, than children presented with the cue cards but not trained in their use (Camparo, Wagner and Saywitz, 2001; Saywitz and Snyder, 1996; Saywitz, Snyder and Lamphear, 1996). Brown and Pipe (2003) and Bowen and Howie (2002) replicated the positive findings of Saywitz and colleagues when they investigated the effects of NE along with the mental reinstatement of context (from the CI) and after a delay. Unfortunately, the NE technique has not been studied in forensic settings, where

the timing of the cues relative to the disclosure of relevant information could affect their suggestiveness and thus their suitability for use with child witnesses.

Saywitz and colleagues have also developed the Comprehension Monitoring (CM) procedure (e.g. Saywitz, Nathanson, Snyder and Lamphear, 1993) which enhances children's ability to detect and cope with non comprehension through instruction and preparation. As described previously, a communication breakdown between a child witness and an interviewer is not uncommon, and children rarely ask for clarification when they do not understand questions posed to them. The CM procedure enhances children's ability to detect and cope with non comprehension through instruction and preparation. Results have demonstrated that children who receive CM training show a significantly greater increase in accuracy than control groups (Saywitz Nathanson, Snyder and Lamphear, 1993). Findings have also shown that CM is a skill which improves with age (Saywitz and Snyder, 1993). As with the NE technique, CM has not been studied in a forensic context.

2.3.3 NON-VERBAL PROPS AND PROMPTS

"There are limits to what children under seven can convey to others using words alone" (Garbarino and Stott, 1992 p.67). Research has shown that the use of external cues may augment children's responses to open questions by serving as aids to the retrieval and reporting of information compared to standard verbal interviews. Younger children in particular might be expected to benefit from such cues given their limited ability to search their memories spontaneously and construct a free description (Pipe, Gee and Wilson, 1993). Relying on the reduced demands of cues as a recognition task also compensates for immature language development by reducing the language demands of a recall task; "Allowing the child to 'show' or 'draw' as well as to 'tell' reduces reliance on language ability" (Pipe Salmon and Priestley, 2002, p. 161). Cues may also

reduce the social and emotional demands of the interview. The introduction of a prop into an interview context can have a dramatic effect on the process of joint meaning making by serving as an effective joint referent (Westcott and Littleton, in press). In this way, props may limit “the intrusion of the interviewer into the child’s world” (Steward and Steward, 1996, p. 33).

For the purpose of this thesis the following distinction between cues has been made:

- *Non-verbal props* - these are situationally specific retrieval cues, they are representations of actual props present at the event being recalled or can be associated with the alleged offence. Non-verbal props include anatomically correct dolls, toys, scale models and photographs.
- *Non-verbal prompts* are not situationally specific retrieval cues, they are memory aids that are not exact copies of previously seen objects at the event of being recalled.

Non-Verbal Props

Toys, Dolls and Scale Models

Relatively few empirical studies have examined the effects of different kinds of toys, dolls and scale models on children’s reports. However, it is clear from the few studies that do exist, that the extent to which toys facilitate children’s reports is influenced significantly by the age of the child, the type of prop provided and the way in which they are used in the interview (Pipe *et al.*, 2002). Wilkins *et al.* (1991) tested very young children for their recall of an episode in which a clown came into their nursery. They found that verbal prompts led to improvements in the amount of accurate recall compared with free recall. However, when the children were given a model of the nursery, accurate recall increased by 200%.

In a legal context, anatomically detailed dolls are most likely to be considered as props when sexual abuse is suspected. The discussion about the usefulness of toys and dolls in particular, has centred on the concern that these dolls do not reliably discriminate the play behaviour of abused and non-abused children (Everson and Boat, 2002). Additionally, when used with children under 5-years, the use of dolls can fail to increase the quantity of correct information children recalled and actually increases errors (see review by DeLoache, 1995).

Photographs

The use of photographs as a non-verbal prop could potentially minimise the risk of introducing errors as children do not interact with the objects themselves. Children as young as two and a half years understand the relationship between a photo and the items depicted, and this symbolic understanding appears to develop earlier for pictures than for models and toys (DeLoache and Burns, 1993). The majority of recent research studies have focused on which stage of the interview the photograph should be introduced. Salmon and Irvine (2002) compared the effectiveness of a photograph reminder presented at different time points between an event and an interview as a means of enhancing recall. They found that a photograph reminder presented 24 hours before the interview enhances children's recall to a greater extent than reminders presented at other points. Paterson and Bull found photographs to be effective in aiding recall after a long delay. Five and 6-year old children's verbal recall of a magic show was substantially assisted one year later by showing them a few photographs of objects used in the event (B. Paterson, personal communication, 24th May 2002). These preliminary studies suggest that this is a promising area of research and that photos can be effective retrieval cues following short and long delays. Further work is needed

regarding the effects of distracter photos or of distracter items in photos on accuracy and whether photos are more likely to assist recall of some events than others.

It has been argued that props could be distracting for a child or encourage fantasy based accounts (Ceci and Bruck, 1993). Additionally, the effectiveness of props is influenced by the child's ability to detect and respond to the correspondence between the prop and its referent: the child must understand that the prop represents something that was present during the event. However, research has shown that most children can understand the symbolic relation between a photograph and its referent at a much younger age (around 36–42 months) than for models and dolls, even with relatively low levels of similarity between the two (DeLoache and Burns, 1993).

Pipe *et al.* (1993) investigated the effects of various cues on children's recall of a brief interaction with an adult. They found that the props used did not need to be exact replicas of the original objects. Props which were not identical, but were related to the originally experienced objects (i.e. non-verbal prompts) improved recall and did not lead to a decrease in accuracy.

Non-Verbal Prompts

A number of research studies have investigated the use of non-verbal prompts as retrieval cues. Researchers propose that the use of prompts which are not situationally specific mean that there is less chance that recall will become contaminated through the provision of specific information, and would therefore, reduce the potential for over-elaboration noted in the studies described above.

Ling and Blades (2000) investigated the provision of a non-verbal memory aid. They used a colour chart which could be used in situations in which memory for colour is

important, such as in forensic interviews (descriptions of criminals usually mention clothing, skin and hair colour). Results support those of previous researchers who found that non-verbal cues improved the recall of children (e.g. O'Callaghan and D'Arcy, 1989). The advantage of the colour chart not being situationally specific meant that in contrast to previous findings, the prompts did not decrease the accuracy of children's recall when compared to that of those in the control condition. They found that even very young children can make use of non-verbal prompts to facilitate their recall.

In a study conducted by Aldridge, Lamb, Sternberg, Orbach, Esplin and Bowler (2004), 4- to 13-year-old alleged victims of sexual abuse were interviewed by police officers using an investigative interview protocol, following which they were shown a human figure drawing and asked a series of questions. Again, results showed that the use of the drawing helped elicit new forensically relevant details, and the drawing was especially useful with the younger children (4- to 7-year-olds). However, unlike the findings of Ling and Blades (2000), the additional information elicited was associated with higher error rates. This was attributed to the fact that the interviewers were relying more on focused recognition memory prompts. Therefore, Aldridge *et al.* (2004) concluded that drawings should only be introduced as late as possible in investigative interviews in order to minimise contamination.

Importantly, Aldridge *et al.* (2004) took a number of steps to minimise the risk of suggestibility, including, for example, using a gender-neutral outline drawing rather than an anatomically detailed drawing, reasoning that explicit drawings might be too suggestive. Additionally, the drawing was only introduced after the interviewers felt they had elicited as much information as possible from the children using the structured

interview protocol. This was to ensure that any contamination was minimised and the forensic value of the information elicited earlier in the interview was not compromised.

Drawing

It has now been repeatedly demonstrated that asking children to draw during verbal interviews can also facilitate children's verbal recall. Gross and Hayne (1999) assessed 5- to 6-year-old children's recall of a novel event after a number of different delays up to one year. Across all delays, children who were given the opportunity to draw verbally reported more information than children merely asked to tell. Importantly, the increase in recall did not occur at the expense of the children's accuracy. However, other recent research suggests that the 'Draw and Tell' interview technique can actually promote false reporting. Strange, Garry and Sutherland (2003) showed that being encouraged to draw offered the opportunity to create fantasy details which may continue into when children are questioned about other events. Another consideration when interviewing children is that drawing extends the duration of the interview relative to a verbal interview.

2.3.4 TRAINING OF INTERVIEWERS

One of the most important shifts in thinking about children as witnesses has been the removal of responsibility from the child alone for the provision of accurate information, to a sharing of the responsibility with the adult interviewer. The role of the interviewer is "complex, challenging and multifaceted" (Baker-Ward, 2002, p.32). There are currently a number of training courses available to interviewers which provide specific training for conducting interviews, such as CI, PEACE or ABE training. These courses cover both adult and child witnesses. There are also numerous findings which have been published which inform practitioners (such as police investigators or social workers)

about techniques to use when interviewing young children (e.g. Wilson and Powell, 2001).

Although there is a wealth of research on appropriate interview techniques for use with children, agreement about how interviews should be conducted has had little impact on the quality of investigative interviews conducted in England. In an examination of 40 interviews since the implementation of the MoGP (Home Office, 1992), Davies, Wilson, Mitchell and Milsom (1995) report that the standard of the interviews was very good. However, two main difficulties of interviewers were highlighted: allowing children to provide a free narrative without interrupting; and, closing an interview in an appropriate manner. More recently, Sternberg, Lamb, Davies and Westcott (2001) evaluated the quality of interviews using over 100 transcripts of alleged victims between the 4- and 13-years. Raters classified the types of prompts used by the investigators. They found that forensic interviewers in England and Wales relied heavily on option-posing questions and seldom used open questions to elicit information from the children. They concluded that despite the clarity and specificity of the MoGP, its implementation appears to have had less effect on the practices of forensic interviewers in the field than was hoped.

The results of a study by Westcott and Kynan (2004) support and extend this earlier research. In their study, 70 videotaped MoGP interviews from England and Wales, with children aged from under 7- to 12-years of age, were transcribed and coded using a specifically developed scheme. Rapport, closure and free narrative phases varied in the degree to which they were present, and in their quality even when included, and results suggest a number of areas require attention, including the importance of interviewer training.

2.4 CHAPTER 2 SUMMARY

This chapter provided a developmental backdrop against which children's abilities as witnesses should be judged. In summary, the research studies cited demonstrated that children's memory, language and communication abilities improve with age and a number of age heuristics were provided. However, it was also argued that there is no simple relationship between a child's age and his/her ability as a witness, and that considerable individual differences exist. In terms of composite construction, the literature does not offer clear guidance on the youngest age at which a child may be able to construct a composite, and suggests that age itself might not be a good predictor of how well a child might communicate the face of a perpetrator to a composite operator.

Cognitive and social considerations, including research on the development of memory, language and communication, suggested that there are limits on the extent to which children can verbalise their experiences. The chapter went on to consider research which has used 'special' interview techniques such as non-verbal props and prompts to elicit descriptions from children and to enhance the communication process between an interviewer and a child witness. Non-verbal prompts in particular were illustrated as an appropriate interview technique to facilitate even very young children's recall. The chapter described how the use of non-verbal prompts, which are not situationally specific, reduce the potential for recall which is contaminated through the provision of specific information and over elaboration which is sometimes associated with non-verbal props. The exploration of the application of such non-verbal prompts to assist children with verbal facial descriptions and subsequent composite production was one of the central aims of this thesis.

The following literature review chapters consider children's general abilities as witnesses (discussed in this chapter) in relation to children's production of verbal descriptions and composite production of unfamiliar faces.

CHAPTER 3:

FACIAL COMPOSITE PRODUCTION: DEVELOPMENTAL ISSUES

Before the practical issues of facial descriptions, recall and recognition in terms of composite systems are addressed (Chapter 4) it is necessary to consider each of these from a theoretical perspective. This chapter begins with a brief discussion of the distinction between recalling and recognising faces. The discussion then proceeds to the development of face processing. Children's performance at facial recognition tasks and research which has focused on the recognition of unfamiliar faces is considered next. The last section of this chapter considers facial recall in detail, and in particular, verbal descriptions of unfamiliar faces.

3.1 RECALLING AND RECOGNISING FACES

There are different methods of accessing a memory of a face which are based on different skills. When witnesses are working to create a visual representation of a suspect's face, for example to produce an E-FIT (this process is described in Chapter 4) they are required to *describe, recall* and *recognise* the face of the perpetrator (Pike, *et al.*, 2000). The majority of existing research concerned with memory for faces has been directed towards facial recognition and facial recall has received much less attention. Laughery, Duval and Wogalter (1986) state that this is due to the fact that in everyday tasks facial recall or facial descriptions are much less common than facial recognition, thus research on the recognition of faces is more ecologically valid. "We have to recall a person's face far less often than to recognise it" (Shepherd and Ellis, 1992, p.87). However, this should not exclude facial descriptions from research. Facial recall is of

considerable importance and the ability of a witness to provide a clear and detailed description of a suspect can greatly assist a case.

As both facial recall and facial recognition are aspects of memory for faces, to some extent they would be expected to share common characteristics. The question of how recall and recognition are related is one of the oldest in the study of memory (Baddeley, 1990). Research in verbal and pictorial memory, as well as in face memory has shown that, although there is some overlap between recall and recognition, there are important differences between these processes that make generalisations of the findings from one area to the other difficult. In terms of the production of a facial composite, Turner (2004) stated that producing a facial composite from memory is a somewhat different task to that of recognising a presented face. It is therefore uncertain to what extent knowledge derived from research into face recognition is applicable to facial compositing. This is further explained by the fact that during recall a witness is required to break down a face bit by bit so that information on different features or areas of the face can be recalled. In contrast, the facial recognition process is a holistic matching task (Cohen, 1989). The development of face processing will be considered next before the discussion of recognition and recall is returned to.

3.2 THE DEVELOPMENT OF FACE PROCESSING

Face processing has been a great point of interest in developmental psychology for a number of decades. The majority of analyses of the development of face processing have focused on the *encoding* of information from a face. Facial processing in infancy has been explored in much detail. Key findings are that infants will track faces and that infants prefer to view a schematic face rather than a blank head or scrambled face (e.g. Johnson, Dziurawiec, Ellis and Morton, 1991). However, due to the focus of this thesis

on child witnesses and stranger perpetrated crime, the current discussion will be restricted to the development of facial processing of *unfamiliar* faces in children aged 4-years-old and upwards. Research that is relevant to the representation of familiar faces will not be considered except where this topic clarifies the main direction of the chapter.

Using simple recognition tasks, empirical evidence indicates that young children have more difficulty than adults when encoding and subsequently identifying unfamiliar faces and that children's facial processing skills improve within the first decade of life (for a review see Chung and Thompson, 1995). There has been a long standing debate in the literature concerning whether, and to what extent, adults and children differ in the way in which they encode information about unfamiliar faces. Brigham (2002) described the distinction between the two main perspectives on developmental changes in encoding: (1) is there a *qualitative* change in processing information - are children of different ages and adults using different encoding processes? or (2) are there *quantitative* changes in processing information - are older children simply encoding more information? Research which has attempted to address these questions will be described below. However, it is necessary to first provide some definitions of facial information.

3.2.1 DEFINITIONS OF FACIAL INFORMATION

There is an important and useful distinction in the literature between *featural* and *configural* information contained within a face (e.g. Rakover, 2002). Featural information relates to facial features which can be referred to in relative isolation (e.g. the size and shape of the eyes, nose and mouth). Configural information refers to the spatial relation between the features. These sources of information are not independent of each other.

Early researchers (e.g. Carey and Diamond, 1977) argued that younger children encode and recognise faces via a featural encoding strategy that focuses on isolated single, salient features of faces. In contrast, older children and adults utilise a configural encoding strategy of information relative to the global appearance of a face. This configural processing takes into consideration the relationship between various facial features. For featural encoding, as age and experience increase children are able to scan faces in a more systematic and organised manner, thereby encoding additional relevant features. Research (e.g. Carey and Diamond, 1994) has investigated the development of configural information in face perception to determine whether young children are deficient in processing configural information of unfamiliar faces when compared to older children and adults. This sensitivity to configural information is often measured indirectly by the size of the inversion effect (Leder and Bruce, 2000) and a major stream of research has investigated inversion effects and their interpretation in terms of configural versus featural processing.

Generally, faces are recognised more easily than any other class of stimuli that are as similar to one another in their configuration. However, by inverting them (turning them upside down) faces become harder to distinguish than other classes of inverted stimuli. This effect was first reported in a number of studies by Yin (e.g. Yin, 1969). Yin argued that the unique reversal of recognition accuracy for faces, from best upright to worst inverted for face recognition was the product of a system distinct from the one used for recognising other types of visual stimuli. The composite task of Young, Hellawell and Hay (1987) and the part-whole paradigm developed by Tanaka and Farah (1993) also confirmed this interpretation. Research which has employed these and other methodologies (including paraphernalia studies) for studying the development of

configural processing in children is discussed below, before the definitions of facial information are returned to.

3.2.2 FROM EARLY TO CONTEMPORARY RESEARCH

Goldstein and Chance (1964) carried out the earliest systematic studies of the development of face recognition abilities. They showed that the performance of children aged between 5- and 13-years improved with age on a recognition memory task. Over a decade later, Carey and Diamond (1977) conducted a series of experiments, and showed 6-, 8- and 10-year-old children upright and inverted photographs of unfamiliar people. Immediately after viewing the photographs, children were asked to identify the previously seen face from a face they had not viewed. Their results revealed that 6- and 8-year-olds encoded upright faces as poorly as inverted faces however, by age 10 a significant effect of orientation emerged. In a second experiment Carey and Diamond (1977) tested children's identification of faces across changes in paraphernalia (e.g. Clothing, Hair Style and Eyeglasses) and facial expression. Their prediction was that, if young children attend more to isolated featural information in faces, they would be more likely than older children to make false recognition decisions based on paraphernalia and facial expression. Consistent with their prediction, they found that the 6-year-olds were misled more by changes in paraphernalia and facial expressions than 8- and 10-year-olds.

Drawing on this evidence from their inversion and paraphernalia studies, Carey and Diamond proposed that the increasing accuracy in face recognition abilities during childhood may reflect a qualitative shift in processing modes. They claimed that until the age of 10-years, children rely on featural based processing to recognise faces and

that there is a *switch* from this featural encoding to configural processing at age 10-years to make face recognition judgements.

Since this early research, other researchers have challenged Carey and Diamond's (1977) encoding switch claims (e.g. Flin, 1985; Carey and Diamond, 1994). Flin argued that the recognition task used in Carey and Diamond's study may have been too demanding for the 6-year-olds and that their data were undermined by floor effects. Even though children responded above chance levels, their generally poor performance might have obscured possible inversion effects. In contrast to Carey and Diamond (1977), Flin found that the magnitude of the differences in the upright and inverted conditions was the same for the 6-year-olds as for the 8- and 10-year-olds. Flin argued that, when age-related performance differences are taken into account, there is little evidence to support the encoding switch hypothesis.

Carey and Diamond (1994) studied face processing in 6- to 10-year-olds using the composite task developed by Young *et al.* (1987). Children studied a set of faces made up from the upper and lower halves of different faces. Children were asked to name the upper half of a face (while the lower half of the face belonged to a different face). In different conditions, the face halves were aligned (composite faces) or not aligned (non-composite faces) and the faces were shown upright or inverted. Carey and Diamond (1994) found 6- to 10-year-old children perceived and processed the upper half of the face configurally, in conjunction with the lower half. They found that the inversion effect was inversely related to age, where 6-year-olds were less affected by inversion than were 10-year-olds who were less affected than were adults. They conclude that there is an increased reliance on configural information between these ages, so 6-year-old children do process faces configurally, but only to a limited extent. Therefore Carey

and Diamond moved away from proposing a *qualitative* change in processing to a *quantitative* change in processing with age.

More recent research (Friere and Lee, 2001; Mondloch, Geldart, Maurer and LeGrand, 2002; Tanaka, Kay, Grinnell, Stansfield, and Szechter, 1998) also argued against the encoding shift between ages 6- and 10-years and provided evidence indicating that children process faces in a manner that is qualitatively similar to the way that adults process faces. This research suggests that, at all ages, children encode faces in terms of configural features and are sensitive to the orientation of the face. A summary of each of these studies is provided below:

Tanaka *et al.* (1998) used the part-whole paradigm (Tanaka and Farah, 1993) in which recognition of isolated face parts (Eyes, Nose and Mouth) is contrasted with recognition of the same features represented in a whole face context. They asked 6-, 8-, and 10-year old children to learn a set of unfamiliar faces along with their respective names. Children were then asked to recognise single facial features or features within a whole face context. Tanaka *et al.* reasoned that if children represent faces configurally, then they should recognise a facial feature better when it is embedded in the whole face than when it is presented alone. Results showed that children were more successful at recognising features in the context of a whole upright face. However, for inverted faces, this advantage disappeared, and recognition performance was similar in the whole and single conditions. In conclusion, Tanaka *et al.* made two points about face recognition development: first, that even from age 6-years children process faces configurally, that they are sensitive to a change in facial configuration; secondly, the configural advantage remains relatively stable from age 6 to age 10 and is maintained through adulthood.

Friere and Lee (2001) set a target face amongst distractors that varied along configural dimensions in their first experiment or feature dimensions in their second experiment for 4- to 11-year-olds. Results indicated that even with a brief exposure to a target face the youngest children (aged 4-years) processed both featural and configural information in order to make identity judgements. However, these children found configural processing more difficult than the processing of facial features and were slower in the configural task.

Finally, Mondloch *et al.* (2002) used a set of faces varying on features, external contour or internal spacing of features with 6-, 8- and 10-year-olds. Their results showed an improvement in performance for all age groups in all conditions. However, the improvement was less marked in the feature and external contour conditions compared with the internal spacing condition. Mondloch *et al.* interpreted these results as showing that the processing of features and external contours approaches (but does not quite reach) adult levels in children as young as 6-years. The configural processing (required most in the internal spacing condition) develops more slowly than the processing of component information until the age of 10. So that, although 6-year-old children use both featural and configural information, it is not until 10-years of age that they have become increasingly more reliant on using configural information to distinguish between faces.

Overall, these studies demonstrated that children aged 6-years and older process faces in a manner that is qualitatively similar to the way adults process faces⁹. Consistent with the most recent study by Mondloch *et al.*, all of the research described above suggests that configural processing develops at a slower rate than featural processing,

⁹ Research has also shown this is also true for younger children aged 4- to 5- years (e.g. Pellicano and Rhodes, 2003).

highlighting that configural processing is weaker in young children compared with older children and adults.

As illustrated by the studies cited above, children and adults may use both kinds of facial information - featural and configural - depending on the requirements of the task and its degree of difficulty. Flin (1985) argued that Carey and Diamond's (1977) findings might have been due to the task they used. Indeed, the majority of studies examining face processing in children use adult style learning tasks, and therefore suffer from floor effects in the younger age groups. Under these conditions the presence or absence of the desired effect cannot be properly assessed (Gilchrist and McKone, 2003). New tasks are needed to look at young children's facial processing. Usefully, Bruce, Campbell, Doherty-Sneddon, Import, Langton, McAuley and Wright (2000) have published a set of face processing tests suitable for use with children aged from 4 to 10-years. In contrast to the usual test of face identification skills, these tests examine a wider range of face processing abilities. The tests are developmentally sensitive and include expression tests, facial speech and sound tests, gaze tests, and identity matching tests. Bruce *et al.* stated that their battery of tests should prove useful to identify children whose face processing skills are not in line with their age group.

Gilchrist and McKone (2003) make another methodological point: that even if performance is off floor level, the size of effect across an age group cannot be compared in quantitative terms unless some baseline condition is equated across groups. For example, Carey and Diamond (1994) and Mondloch *et al.* (2002) claim evidence for development in strength of the phenomenon of interest across ages, but ignore the presence of overall difference in performance levels with age.

A further criticism of earlier research is related to the definitions of configural facial information as described in section 3.2.1. As research has progressed so has the understanding of configural processing and there are now multiple definitions. A number of research studies (e.g. Pellicano and Rhodes, 2003; and Maurer *et al.*, 2002) distinguish between three types of configural processing: (1) sensitivity to first-order relations (seeing the arrangement of features within a face: two Eyes above a Nose, which is above a Mouth); (2) holistic processing (representations of the overall structure of the face); (3) sensitivity to second order relations (representations of the spatial relations among the individual features). However, many research studies use the term 'configural processing' to refer to one, two or all three of these types of configural processing.

In a review of the literature, Chung and Thompson (1995) explain the developmental differences in facial processing in terms of the expert thesis (e.g. Carey and Diamond, 1994; and Diamond and Carey, 1986). This distinguishes between two distinct sources of configural information in a face. One kind of configural information (first order) is perceptual, innate, automatic and accessible from early on in life. Whereas second order configural information is a process of the cognitive recognition of the spatial relations among facial features. This kind of information develops with age as it is acquired through experience.

The most recent research (e.g. Donnelly and Hadwin, 2003; Gilchrist and McKone, 2003) takes into account all of the methodological criticisms of earlier studies. These studies highlight the importance of task demands; employ tasks that avoid floor and ceiling effects and require a higher level of sophistication in their measurements; and, also use more detailed descriptions of configural processing. Importantly, the findings

from these studies still provide support for the notion that children process faces in a manner qualitatively similar to adults

Donnelly and Hadwin (2003) reported the ability of young children to detect the Thatcher illusion¹⁰, where the effect is only present when the face is upright and not when the face is inverted. The presence of the illusion indicates the computation of configural information between individual features. In their first experiment Donnelly and Hadwin showed that the children could perceive the Thatcher illusion when they were as young as 6-years. However, some performances reached ceiling level. In a second experiment, by changing the stimuli used to assess more effectively the emergence of configural face processing, they showed this only holds under specific conditions. The results illustrated that although configural processing exists in younger children it is not equivalent to that found in adults or older children. Overall, the results from Donnelly and Hadwin's two experiments demonstrated that where ceiling performance is avoided across all age group, children as young as 6-years can process configurally and that configural processing becomes more resilient with age.

Finally, another recently published research paper, Gilchrist and McKone (2003) provide a review of studies which have examined face processing in children. Studies which have examined inversion and face space effects were included in the review along with a new experiment which investigated distinctiveness effects on memory. Gilchrist and McKone (2003) concluded that each of the major aspects of adult style face processing (inversion effect, first and second order relational processing, race effects, distinctive effect) are present by 6- to 7-years of age. In conclusion, Gilchrist and McKone argued that the overall improvement in speed and accuracy beyond this

¹⁰ In the Thatcher illusion, the eyes and mouth are inverted with the effect that the resultant face appears grotesque.

age reflects a development in general cognitive abilities rather than any changes in face processing per se.

Research suggests that the manner of encoding facial information does not change with age; that there is a quantitative change rather than a qualitative change. Rather than viewing younger children as encoding a face on the basis of its features and then switching to a more configural encoding strategy, the accumulated findings of research indicate that by the age of 6-years children are sensitive to changes in facial configuration and are perceiving and remembering faces holistically. However, the development of configural processing is weaker in such younger children. Most importantly, contemporary research has found that children (even young children) are considerably more competent in processing faces than frequently believed.

Although there is a broad consensus in the literature about an improvement in facial processing skills with age, there is no consensus concerning the age by which development is complete (Flin, 1980). The age at which configural processing is fully developed is still debatable. Some studies have shown a temporary decline in face processing. Chung and Thompson (1995) provided an extensive review of the literature and cited 12 different studies which used a variety of participants, stimulus faces, recognition measures and testing procedures. These studies demonstrated a steady development in processing of unfamiliar faces with age. However, a developmental discontinuity, or 'developmental dip', was also observed around the age of 12-years. The magnitude of the dip is not always significant and it has also been observed at

different ages. This inconsistency in the research findings raises the question of whether the occurrence of a developmental dip is a genuine and reliable phenomenon¹¹.

In conclusion, it is important to note that such explanations of facial processing described above may have failed to account for the *complete* development of facial processing. By manipulating the stimuli to be encoded the studies described only consider the initial stage of encoding as the sole source of developmental change. This is only of limited application to composite construction where all aspects of facial processing are used (Pike *et al.*, 2000). Perhaps further research examining *all* aspects of facial processing would be useful for application to the task of composite construction. For example, it would be helpful to understand how developmental changes may influence how faces are stored as mental representations and how developmental changes may impact upon recall as opposed to recognition. Current computerised composite systems involve both recall and recognition processes however there are new systems in development that could draw largely on recognition processes¹².

Children's facial recognition is considered next, before the discussion turns to facial recall.

3.3 FACIAL RECOGNITION

3.3.1 CHILDREN'S RECOGNITION PERFORMANCE

Both adults' and children's performances on facial recognition tasks are well documented. Chance and Goldstein (1984) reviewed the estimated accuracy of face

¹¹ This thesis is concerned with children under the age of 10 years and therefore is not concerned with any developmental discrepancies which occur over the age of 12.

¹² These newer composite systems are introduced in Section 4.1.2.

recognition memory performance of children and concluded that the number of correct identifications appears to increase from around performance levels of 35-40% for 4- to 5- year olds, 50 to 60% for 6- to 8-year-olds, 60 to 70 % for 9- to 11-year-olds and 70 to 80% for 12-year-olds. Adult performance has been found to be similar to that of 12- to 14-year-old children.

Pozzulo and Lindsay (1998) examined the identification accuracy of children and adults in a meta analysis. Overall, results showed that very young children (4-year-olds) were less likely than adults to make correct identifications. However, when shown a target present line-up, children aged 5-years and over did not differ significantly from adults with regard to correct identification rate. The section below describes the recognition of unfamiliar faces in more detail in an attempt to explain the reasons for these developmental differences.

3.3.2 RECOGNITION OF FACES (INNER VERSUS OUTER PARTS)

Research studies have demonstrated a developmental trend in the recognition of *familiar* faces: where young children's representations are dominated by the outer features of the face (the Hairline, Chin and Ears) and adults representations are dominated by inner facial features (the Eyes, Nose and Mouth) (e.g. Campbell, Coleman, Walker, Benson, Wallace, Michelotti and Baron-Cohen, 1999). As previously described, the development of facial processing abilities also involves the processing of the configuration of features. The inner part of the face contains the majority of the elements that are involved in configural processing. The effect of familiarity on recognition from inner and outer parts of faces therefore involves configural processing. The research described in Section 3.2.2 suggests that although young children process faces configurally this skill is weaker than in adults, and may provide an explanation for

why children find it harder to recognise people from their inner, rather than their outer features. As children develop and rely more on configural processing, they become better at processing the inner features of faces and eventually recognition based on the inner features overtakes recognition based on the outer features.

In contrast to the research on familiar facial recognition, research studies have shown that adults' representations of *unfamiliar* faces are dominated by the outer features of a face (the Hairline, Chin and Ears) rather than the inner facial features (the Eyes, Nose and Mouth) (e.g. Young, Hay, McWeeny, Flude and Ellis, 1985). In terms of children's representations, little is known about the developmental pattern of importance of inner and outer facial features for recognising unfamiliar faces.

Want, Pascalis, Coleman and Blades (2003) conducted a study which describes the pattern of unfamiliar face recognition. Children aged 5-, 7- and 9-years were presented with the task of recognising an experimentally familiarised face (initially seen in a short video) from a still picture of the whole face, just the inner, or just the outer parts. The results showed that, for all ages, recognition was faster (and in most cases, more accurate) for outer features alone than for inner features alone (and was fastest for whole faces). Want *et al.* concluded that their findings demonstrate the importance of outer facial features for the recognition of relatively unfamiliar faces.

3.4 VERBAL DESCRIPTIONS OF FACES

As noted above, in comparison with the heavily researched area of face recognition, facial recall has received much less attention. It was not until a paper by Ellis, Shepherd and Davies (1975) that the recall of faces was studied. Before this, there was not a suitable instrument to test participants on their ability to reconstruct their recollection of

a face, the only method available was a recognition procedure. Ellis (1986) stated that the intrinsic difficulties of verbal descriptions led to facial composite systems being developed as a practical alternative. The introduction of composite kits provided a suitable instrument with which to investigate participants' abilities to recall a face¹³.

As already stated, facial recognition cannot be easily compared to facial recall. Laboratory experiments report low levels of accuracy for facial recall, lower than for recognition (Prior, 1996). Shepherd and Ellis (1992) suggested that our ability to recall faces may be poorer for many reasons. They suggest that it may be that people can easily form a mental image of someone they have seen but simply find it difficult to externalise that image in a form that others can utilise.

3.4.1 FEATURE SALIENCY

Feature saliency research has focused on which facial features are encoded, and on any change in the saliency of these features in the representation of the face. From studies involving adults it is clear that not all features of a face are remembered equally well and some features (salient features) are the basis for subsequent recognition (Shepherd, Davies and Ellis, 1981)¹⁴.

There is little evidence available on how children process individual facial features (for a review see Chung and Thompson, 1995). Additionally, studies concerning the age at which children attain patterns of feature saliency are hindered by the lack of consensus over what constitutes an adult pattern of feature saliency (Prior, 1996). Flin (1983)

¹³ Facial recall with regards to composite systems will be discussed in Chapter 4.

¹⁴ There is also some evidence that feature saliency may differ for familiar and unfamiliar faces (Ellis, Shepherd and Davies, 1979).

reported that on a facial composite task using the Photo-FIT system¹⁵ 7-year-olds and adults showed the same feature saliency. Pedelty, Levine and Shevall (1985) also observed no age difference in the pattern of salience with 7-, 9- and 12-year-old children and adults. It has been argued by Chung and Thompson (1995) that these results should be discounted for using artificial stimuli.

Chung (1991) used a standard task and a more realistic type of stimulus, and again observed no age differences between children aged 7- and 11-years and adults. In a forensically realistic study, Davies, Tarrant and Flin (1989) explored the accuracy and completeness of children's testimony for a range of event and appearance information. Children aged 6- to 11-years participated in a simulated health check procedure. For all children, Hair was the most dominant characteristic recalled. Apart from Eyes, few dominant facial features were mentioned by the youngest children. The most commonly recalled features for 10- to 11-year-olds were Hair, Freckles/Spots, and Nose. Overall, the order of saliency was similar to adults (especially for older children). A more recent study by Prior (1996) supports these findings. Prior provides the following generalisation of the adult pattern of feature salience with unfamiliar faces with Hair > Eyes = Mouth > Chin > Nose and reports that by 10- to 11-years of age children's patterns of feature salience are almost identical to adults.

3.4.2 TERMS FOR DESCRIBING FACES

The vocabulary for describing faces has frequently been described as insufficiently rich or precise for it to be an effective way to communicate information about faces (Phillips, 1978). "There does not appear to be much linguistic richness in our vocabulary to describe faces" (Laughery *et al.*, 1986, p.385). Facial patterns and

¹⁵ The Photo-FIT composite system is described in Chapter 4.

configurations are extremely difficult to represent in words (Clifford and Bull, 1978) and verbal labels do not exist because stimulus attributes do not lend themselves easily to verbal associations (Bourne, 1966).

In the specific area of facial recall, the majority of research has concentrated on collecting adults' terms for describing faces (Laughery, Duval and Fowler, 1977; Shepherd, Ellis and Davies, 1977). From research with adults, it is known that the adjectives used to describe faces are largely feature independent, size and shape words (Laughery *et al.*, 1986). Research conducted by Shepherd and colleagues in the 1970s and 1980s concentrated on adults' vocabulary for describing faces. In a study by Shepherd and Deregowski (1981) adult participants were asked to write free descriptions of black and white prints of faces. The feature descriptors mentioned were then selected as scales for rating faces and performing a factor analysis. Based on Kelly's Personal Construct Theory (1955) adult participants were then presented with triads of faces. They were asked to select which two faces were similar in one respect to each other and different from the third face and to describe the basis of the differences. These descriptions were used as a basis for a facial feature index in the computerised composite E-FIT facial recall system¹⁶.

The feature index in E-FIT represents around 10 years work by the Psychology Department, University of Aberdeen, Scotland. The phrases used in the index are based entirely upon the common words used by a large population of adult test participants when describing faces in their own terms. The Aberdeen Index is described in the E-FIT help file as "a down to earth and easy to remember system that shrouds a sound

¹⁶ The E-FIT composite system is described in more detail in Chapter 4.

scientific approach to categorising facial features” (E-FIT v3.1a for Windows User Manual, p.59).

E-FIT is set up to consider the face as seven separate sections corresponding to the following individual features: Face Shape; Hair; Eyes; Eyebrows; Nose; Mouth; Ears. Within these main features there are 54 feature elements and a total of 191 exemplars. Based upon the research with adults described above, each feature, element and exemplar has a description of what the term means. An example of the descriptions for the feature element Face Shape (which is a feature element of the main feature Face) and its exemplars is provided below¹⁷:

Feature: FACE

Description: *“The blank background image or “potato” on which the other features are built”*

Element: FACE SHAPE

Description: *“Face shape is defined as general outline of the head in frontal view and is dependent mainly on the skull and jaw line”*

Exemplar: OVAL

Description: *“Face which is long in relation to the width. Smooth jaw line without sudden angle changes. Narrow chin, but not pointed”*

Exemplar: ROUND

Description: *“Similar to oval face, but shorter in relation to width. Smooth jaw line without any sudden angle changes. Wider chin”*

Exemplar: TRIANGULAR

Description: *“Wide across forehead. Pointed chin”*

¹⁷ In addition to Face Shape, Face also has 11 other Elements including Width, Length, Chin Shape and Forehead.

Exemplar: SQUARE

Description: *“Mainly determined by shape of chin - having straightness of line at the base of the chin. Note: Includes rectangular faces”*

Exemplar: ANGULAR

Description: *“Outline of jaw is marked by distinct changes in angle of jawbone and chin”*

However, there are a number of problems with using this system with children. First, the terms and descriptions were devised by adults for adults. At the time of writing there are no comparable descriptions from children. There have been few systematic studies of children's abilities to systematically describe a face and there has not been any research which has looked at children's terms for describing faces at the same level as has been conducted with adults. Yet this information may be essential to enable the police to identify a culprit (Davies, Stevenson-Robb and Flin, 1988). The studies which have been conducted so far concentrate on the completeness, accuracy and order of children's descriptions (e.g. Bull, 2001; Cunningham and Odem, 1986; Pedelty *et al.*, 1985; Prior, 1996). These studies have shown that facial descriptions, like any recall task, are less likely to be complete and/or accurate at very young age levels “however, this is not meant to imply that even very young children...could not provide very useful descriptive details about another person” (Sporer, 1996, p.68).

A second problem with the facial feature index described is that witnesses do not have ‘standard’ descriptions i.e. what one person describes as long hair, may be described as short by another. This results in a reliance on composite operators to translate a witness’ descriptions, based on their own interpretations, to match the index. The E-FIT user

manual states that “new operators should take time to acquaint themselves with the Aberdeen Descriptors, and when entering a description, translate the witness’ description to match the Aberdeen Index” (E-FIT v3.1a for Windows User Manual, p.32). This difference in use of terms may be additionally problematic for child witnesses (as will be illustrated in Chapter 4).

It is common in police practice to find that the spontaneous verbal descriptions provided by a witness tend to be vague, general and incomplete (Davies, 1983). Davies (1986) stated that most police forces will use some form of cued recall or check list procedure for eliciting the appearance of a suspect. These will include facial features and a selection of alternative descriptors to assist the witness and police officer, which are usually based on the research described above with adults. Such checklists are clearly useful, though there is a lack of standardised format within and across forces, and there is no equivalent list for working with child witnesses. The use of such special interview techniques and their effects on the communication process between a witness and an interviewer were described in Chapter 2. Sporer (1996) suggested that future research exploring new techniques such as adjective check lists, rating scales or alternative choice visual displays might prove useful in making the communication process between a witness and a police operator more effective¹⁸.

3.4.3 VERBAL OVERSHADOWING

Finally, there is a phenomenon known as verbal overshadowing which indicates that being encouraged to use language to describe a face will create verbal representations, which may interfere or ‘overshadow’ the ability to subsequently identify that face from

¹⁸ Developing appropriate new interview techniques will be the main driving force for the current programme of research.

a line up (Schooler and Engstler-Schooler, 1990)¹⁹. This effect was first investigated in a series of experiments by Schooler and Engstler-Schooler (1990). Results consistently indicate that participants asked to provide a description of a target person prior to their identification were significantly less accurate in subsequent identification than were individuals who did not attempt a description.

Since this initial set of studies, similar results have been found for facial memory (e.g. Dodson, Johnson and Schooler, 1997). Schooler, Fiore and Brandimonte (1997) proposed three possible main factors that contribute to the verbal overshadowing effect: (1) modality mismatch (the notion of competing verbal vs. non-verbal representations); (2) availability (the assumption that the visual representation remains available in memory); (3) recoding inference (the notion that the overshadowing effects occur because a erroneous verbal trace of the target is created and later recalled).

There have also been a number of studies in which the verbal overshadowing effect has not occurred and some studies have demonstrated the positive effects of verbal description, rehearsal and elaboration on later recognition of faces (e.g. Yu and Geiselman, 1993). The difference in findings may be due to the degree of interference generated by the description task: a meta-analysis by Meissner and Brigham (2001) across 15 research articles found that instructions that asked people to provide very full verbal descriptions are most likely to produce the verbal overshadowing effect. The majority of this research has been conducted with adults, and it remains to be seen if the process occurs in the same way for children (Brigham, 2002). To date there are two studies which have addressed this issue. In a study conducted by Memon and Rose (2002) children aged 8- to 9-years witnessed a live event. After a 24 hour delay the

¹⁹ The effects of composite production of subsequent identification performance of an eyewitness are considered in Section 4.3.3.

effect of a description on an identification parade was examined. The results of this study revealed no verbal overshadowing effect and this result has also been supported in a study by Clifford (2003).

3.5 CHAPTER 3 SUMMARY

This chapter considered the developmental issues surrounding facial descriptions, recall and recognition. In summary, the research cited illustrated that children's facial processing skills improve within the first decade of life. Collectively, the more recent research indicated a *quantitative* change rather than a qualitative change in children's facial processing. Crucially, the contemporary research discussed demonstrated that young children are considerably more competent in processing faces than was frequently believed and that by the age of 6-years children are sensitive to changes in facial configuration and are perceiving and remembering faces holistically (although this development is weaker in younger children).

The chapter emphasised how, in contrast to facial recognition, children's performance at facial recall has received much less attention and that the majority of research on facial recall in general has concentrated on collecting adults' terms for describing faces. Indeed, as described, the E-FIT composite system was built using terms collected from a population of adult participants. On the basis of the research discussed it is argued that this deficit in knowledge of the language and terms children use to describe unfamiliar faces may exacerbate problems in the communication process between an interviewer and a child as described in Section 2.1.4. It was therefore one of the central aims of this thesis to obtain a more detailed understanding of children's verbal descriptions of unfamiliar faces in order to assist children's production of facial descriptions and subsequent composite constructions.

The following chapter considers the impact of the developmental issues of facial composite construction described in this chapter and children's general abilities as witnesses (described in Chapter 2) on the procedure of obtaining facial composites from children.

CHAPTER 4: FACIAL COMPOSITE PRODUCTION: PRACTICAL ISSUES

After considering children's general abilities as witnesses (Chapter 2) and some of the developmental and theoretical issues of facial recall (Chapter 3), it is necessary to consider the more practical aspects of facial composite construction. This chapter begins with a brief history of the systems used to obtain facial images from witnesses. Both computerised and non-computerised systems are described along with a summary of the research that has evaluated these systems. The structure of facial composite interviews will be explained including the consideration of the communication process between a composite operator and a witness. Empirical research addressing adults' and children's abilities to provide visual representations is discussed and finally, research which has considered the effect of composite construction on later identification performance is introduced.

4.1 COMPOSITE SYSTEMS

4.1.1 WHAT IS A FACIAL COMPOSITE?

A facial composite (as used by the police) is a facial image built up from the combination of individually described component facial parts. The purpose of a facial composite is to provide a pictorial *type likeness*, or general resemblance (not an exact photographic likeness) of a suspect's face. This likeness can then be circulated to other police forces and/or to the general public to enable someone familiar with the suspect to recognise the composite, or to enable someone who has come into contact with the suspect to identify them.

Over the years various systems have been used to produce facial composite images and non-computerised systems have been replaced by computerised systems. The systems been most widely used by the police in the UK and internationally are described below, along with relevant research evaluating their efficacy.

4.1.2 FACIAL COMPOSITE SYSTEMS

Non-Computerised Composite Systems

Artists' Sketches

The oldest form of visual representation is an artist's sketch, which has been in use by the police since the 19th Century and is still in use in a number of countries today (Taylor, 2000). This procedure involves a skilled artist sketching a suspect's face from a verbal description provided by a witness. The artist and witness may make use of visual prompts that the artist has on file to select instances of the features that are being described (e.g. the Federal Bureau of Investigation (FBI) catalogue)²⁰ in order to produce an initial drawing. The final image produced by an artist is usually an extremely detailed sketch.

Identikit

The 'Identikit' system was developed in America by Dunleavy (1959, 1975). The first version of Identikit (Identikit I), a line drawing system, was issued in 1959. This was replaced in 1975 by a second, photographic version of Identikit (Identikit II). Facial composites produced using Identikit II were constructed by stacking a number of transparent sheets, each with a photographed facial feature printed on it. The Identikit system separated five facial features: Forehead (with Hairstyle); Eyes; Nose; Mouth; Chin. Exemplars for each of these features could be viewed in a visual form in a

²⁰ The FBI catalogue is described in more detail in Section 7.1.1.

handbook. Additionally, features could be changed and any artistic enhancements could be hand-drawn onto a blank transparency overlaid onto the composite.

Photo-FIT

The 'Photo-FIT' (Penry Photo Facial Identification Technique) system was developed by Jacques Penry (Penry, 1970) and in the 1970s Britain's police forces moved from the artists' sketch to the Photo-FIT system. Photo-FIT was a hand assembled composite system similar to Identikit II. It consisted of photographic examples of five features: Hair and Foreheads; Eyes; Noses; Mouths; Chins. A witness was directed by an operator to a visual index of photographs of facial features and each feature selected would be combined within a frame. Features could be changed, moved and any artistic enhancement could be added by using a transparent overlay.

Computerised Composite Systems

Mac-a-Mug

Mac-a-Mug was produced in 1986 by Shaherazam for use on Macintosh computers²¹. It was developed from the line-drawn features of the original Identikit with an expanded library of features which included Hairlines, Eyes and Eyebrow combinations, Noses, Ears, Mouths and Chins. Witnesses worked with an operator to make an initial selection of features which were then synthesised and displayed as a face on screen for further manipulation by the operator according to the witnesses' directions. Each feature is accessed through a printed visual manual (as with Identikit and Photo-FIT) or on the monitor. Additionally, features could be moved, rotated and resized and artistic enhancement could be added both within the system, as well as allowing the exportation of the image to a more sophisticated graphics package.

²¹ Mac-a-Mug was developed as a research tool and has never been used by police.

E-FIT

E-FIT (Electronic Facial Identification Technique) (Aspley Limited, 1993) was developed jointly by the Home Office and Aberdeen University Psychology Department. E-FIT is currently the most commonly used computer based system for constructing facial composites in Britain (Clark, 2000). E-FIT comprises a library of facial features that have been cut out electronically from photographs. As described in Section 3.4.2, E-FIT separated seven facial features: Face Shape; Hair; Eyes; Eyebrows; Nose; Mouth; Ears. Within these main facial features there are 54 'feature elements' and a total of 191 exemplars²². E-FIT consists of a feature index in which every feature, feature element and exemplar has a description of what each term means. Within E-FIT, features can be easily moved around, re-sized and re-toned and artistic enhancements can be made by exporting the image to a computerised graphic "paint" program. The E-FIT composite system will be considered in more detail in the following sections of this chapter.

Finally, a new generation of computerised composite systems that are currently being developed use a combination of Principal Components Analysis modelling and genetic algorithms to direct development of facial likenesses. Faces are modelled in their entirety rather than by the selection of individual facial features (e.g. Evo-FIT and EigenFIT). Evo-FIT is currently under development at Stirling University and aims to create an image of the suspect by 'evolution'. First, a witness is shown a number of faces containing random facial features. The witness selects those faces that are most similar to the target. The selected faces are then mixed together to produce a new set of faces that more closely resemble the target. The witness then selects a few of these faces to be bred together again. The process of selection and breeding continues until an

²² All of the composite systems described also include a variety of accessories (e.g. Spectacles, Facial Lines). These are not considered in the thesis.

acceptable likeness is reached. Hancock (2003) reported that Evo-FIT avoids the problems at the heart of other composite methods by modelling faces in their entirety rather than by the selection of individual facial features²³.

4.1.3 THE EVALUATION OF COMPOSITE SYSTEMS

Much psychological research has focused on the investigation of the efficacy of the composite systems described above. However, the majority of published research has focused on the evaluation of the non-computerised systems of Photo-FIT (e.g. Davies and Christie, 1982; Ellis, Davies and Shepherd, 1978; Ellis *et al.*, 1975; Jenkins and Davies, 1985) and Identikit (e.g. Green and Geiselman, 1989; Laughery *et al.*, 1977; Laughery and Fowler, 1980). Limited research has been conducted into the evaluation of the computerised system Mac-a-Mug (e.g. Cutler, Stocklein, and Penrod, 1988; Koehn and Fisher, 1997; Kovera, Penrod, Pappas, and Thill, 1997) and more recently E-FIT (e.g. Davies, Van der Wilik and Morrison, 2000; Turner, 2004).

In general, studies have found serious limitations of the earlier non-computerised systems, Photo-FIT and Identikit, in terms of the production of accurate representations of faces due to the inadequate flexibility of these systems (Green and Geiselman, 1989). Davies and Christie (1982) identified that, in non-computerised composite systems, the range of features available was limited, not representative and not organised in a way that was easy for the operator to access from a witness's description. For example, the linking together of features (e.g. Eyes and Eyebrows and Forehead and Hair) in these earlier systems restricted the number of choices a witness could make. In contrast, computerised composite systems provide access to a more extensive library of facial features, and can produce a larger number of combinations of facial features. For example

²³ The new generation of systems are in the primary stages of evaluation with adult participants, and therefore will not be considered further in this thesis.

the Mac-a-Mug system can produce 98 times the number of combinations the Photo-FIT system can produce and over 1200 combinations more than Identikit (Cutler *et al.*, 1988).

The older, non-computerised systems also required any additional artistic enhancements that might be required to be drawn onto transparencies which produced an artificial appearance to the composite. This was especially true of Photo-FIT where the features were fitted together (rather than overlaid) therefore producing join-lines. Research has shown that this artificial appearance can affect the recognition of a composite (Ellis *et al.*, 1978). Computerised systems allow for more appropriate artistic enhancement. With E-FIT, the artistic enhancement is carried out in a paint package which removes the artificiality of the earlier composite systems, and makes any enhancements indistinguishable from the original composite. Computerised systems also allow for a better blending of features, which removes any join lines (which were common in earlier systems).

In order to examine the effectiveness of the composite systems under optimal conditions, Davies, Ellis and Shepherd (1978) and Ellis *et al.* (1978) examined participants' abilities to construct Photo-FIT composites either from memory, after a brief presentation (target absent) or in the presence of the target photograph (target present). The results showed no decline in the quality²⁴ of composites made from memory versus those made in the presence of the target. Laughery and Fowler (1980) reported similar findings with the Identikit system. A number of studies have also shown that initial verbal descriptions (both free recall and prompted verbal descriptions) provide a more accurate likeness than the resulting composite faces (Christie and Ellis, 1981; Davies, Milne and Shepherd, 1983). The failure to find an improvement when the

²⁴ Measures of composite quality are described in section 4.3.1.

target was in view and the superior quality of the descriptions in these early experimental studies seem to suggest that failings in the traditional systems could be attributed to limitations in the flexibility systems rather than simply to memory problems of the witness.

Later research studies conducted with the more flexible computerised systems have shown significant target present/target absent effects which cannot be attributed to the failure of the composite system to make recognisable faces. For example, in a forensically realistic study by Koehn and Fisher (1997), participants interacted with a stranger before constructing his face from memory with the assistance of an operator using Mac-a-Mug. The resulting images were then rated as very poor by a set of independent judges. These findings were in contrast to those of Cutler *et al.* (1988) who reported that composites constructed using the Mac-a-Mug system were useful for identification. The critical difference between these two studies was that composites were created from memory (target absent) in the study by Koehn and Fisher and with the target present in the study by Cutler *et al.*

Koehn and Fisher (1997) suggested that witnesses poor performance under memory conditions could be due to the fact that feature driven composite systems (such as Photo-FIT and Mac-a-Mug) work on the basis of the selection of individual features in isolation and with little or no consideration of configural factors. This is incompatible with the configurational processing that is said to characterise facial processing for adults (Young *et al.*, 1987) and more recently for children (Gilchrist and McKone, 2003) (as described in Section 3.2.2).

Davies and Christie (1982) stated that judgements of features are much more accurate when made in the context of a total face bearing a resemblance to the target than in isolation. This was supported by Bruce and Valentine (1988) who found that when features themselves were selected in a featural way and then built up into a total face, this process might be at variance with the process by which faces are perceived and recognised. In their conclusion, Koehn and Fisher (1997) recommended that composite systems which promote holistic processing at retrieval may yield higher quality images. They suggest that the composite construction process should begin by presenting a face shape within which to add and edit features, or by presenting a face that already contains a set of features based on a witness's initial verbal description, as opposed to presenting features in isolation; such as the E-FIT system does.

Unlike the earlier composite systems, which were developed with little input from psychological research, the E-FIT system is based in part on psychological research and utilises a holistic approach by operating within the context of an entire face. As described in Section 3.2.1, holistic processing is a form of configural processing and is related to representations of the overall structure of the face (Pellicano and Rhodes, 2003; Maurer *et al.*, 2002) i.e. considering the face entirely as a whole, including both featural and configural information. The E-FIT instructions emphasise the importance of such a holistic approach: "It is psychologically important to always display a complete face - witnesses find it hard to deal with bits and pieces" (E-FIT v3.1a for Windows on-line help file). This allows for more global manipulations, with more realistic resulting images (Bruce and Young 1998). Turner (2004) provided some supporting evidence for the positive effects of such a holistic approach in laboratory-based experimental studies. When working within the context of a whole face-image, adult participants produced

composites of better quality likenesses than when working featurally within the E-FIT system.

A study by Davies *et al.* (2000) is the only research to date directly comparing composites created using the Photo-FIT system with the E-FIT system. Composites were created of both familiar and unfamiliar faces by adult participants from memory and then with an operator with the target photograph present. Results showed that judges' evaluation of E-FIT composites using a number of measures were no different to their evaluation of Photo-FIT composites. However, E-FIT showed a significant advantage when the participant constructed a composite with the target photograph present. Judges correctly matched these composites correctly 88% of the time. Under these optimal conditions, witnesses can use the greater power of the computerised composite systems to create a realistic likeness with the target present. However, this advantage could not be extended to creating the composite from memory (target absent). These results indicated that the limited value of composites are not strictly due to the composite system itself, as previously thought, but were also a function of the task required of the witness.

Laughery and Fowler (1980) and Koehn and Fisher (1997) stated that in addition to the need for a more configural approach in composite construction, another possible explanation for their disappointing results is the interview process: "the quality of information elicited from witnesses is limited by the interview process.....composite construction may be enhanced by using an appropriate interview strategy in combination with an effective composite generation system" (Koehn and Fisher, 1997, p.211). Furthermore, the earlier research on non-computerised composite systems emphasised the importance of a detailed and accurate verbal description as the key to an

effective likeness (Davies *et al.*, 1986). Given that children have difficulty describing people accurately (as described throughout Chapters 2 and 3), one might expect these difficulties to contribute to problems in the communication process between a witness and an operator (which will be discussed in the following section) which may in turn result in children in general being less effective than adults in making composite likenesses.

Before witnesses' abilities to construct composites are discussed, with a particular emphasis on the importance of the initial interview, the structure of a composite interview needs to be described.

4.2 THE COMPOSITE INTERVIEW

4.2.1 THE 'TYPICAL' COMPOSITE INTERVIEW STRUCTURE

A summary of the 'typical' procedure used to gain a visual representation of a suspect's face from a witness using E-FIT is given below²⁵ based on information from experienced police operators and a police interview trainer regarding the recommended police procedure being taught to E-FIT operators at the time of writing this thesis (A. Parry and C. Charsley, personal communication, 2nd November 2001; Clark, 2000).

Stage 1 - Preliminarily Discussion

The first stage of the interview is a preliminary discussion between the witness and the operator. This is designed both to establish rapport and to gauge the likely quality of information available to the witness in the light of the circumstances in which the suspect was observed (based on the R. v. Turnbull criteria, 1977).

²⁵ All of the composite systems described follow a similar procedure to E-FIT.

Stage 2 - Initial Verbal Description

Next, the witness must recall information about all aspects of the suspect's face from memory, to provide a verbal description of the face to the operator. The quality of a verbal description can be enhanced by using special interview techniques (as described in Chapter 2). This stage is usually carried out before the computer is switched on.

Stage 3 - Prompted Description

The witness may need to respond to questions posed by the operator about the suspect's face. As described in Chapter 3, most police forces will use some form of cued recall or check list procedure for eliciting the appearance of a suspect (Davies, 1986). Such checklists may include facial features and a selection of alternative descriptors to assist the witness and police officer²⁶.

Stage 4 - Entering the Description

The operator will enter the witness's description of the suspect's face into the E-FIT system using E-FIT's description 'check' boxes. Figure 4.1 shows the check boxes for the Face feature as an example.

²⁶ Specific examples of materials used by operators are described in Chapter 5 and Chapter 7.

Shape		
Face Shape ? <input type="radio"/> Oval <input type="radio"/> Round <input type="radio"/> Triangular <input type="radio"/> Square <input type="radio"/> Angular	Chin Shape ? <input type="radio"/> Oval <input type="radio"/> Rounded <input type="radio"/> Pointed <input type="radio"/> Squared <input type="radio"/> Angular	Fleshiness ? <input type="radio"/> Gaunt <input type="radio"/> Lean <input type="radio"/> Average <input type="radio"/> Chubby <input type="radio"/> Flethy
Length ? <input type="radio"/> Short <input type="radio"/> Average <input type="radio"/> Long	Width ? <input type="radio"/> Narrow <input type="radio"/> Average <input type="radio"/> Wide	Forehead ? <input type="radio"/> Low <input type="radio"/> Average <input type="radio"/> High
Dimpled Chin ? <input type="radio"/> No <input type="radio"/> Yes	Age ? <input type="radio"/> Young <input type="radio"/> Adult <input type="radio"/> Middled Aged <input type="radio"/> Old	Heavy Jowls ? <input type="radio"/> No <input type="radio"/> Yes
Double Chin ? <input type="radio"/> No <input type="radio"/> Yes	High Cheekbones ? <input type="radio"/> No <input type="radio"/> Yes	Dimpled Checks ? <input type="radio"/> No <input type="radio"/> Yes

Figure 4.1: Example of E-FIT's description 'check' box for the feature element Face Shape

Once the description has been entered, the E-FIT system uses 'fuzzy logic' to rank its features database in order of closest match to description (for each feature). Thus if hair was described as "short, dark, straight" the database would be sorted with short, dark, straight hair exemplars early in the order and long, blonde, curly hair exemplars late in the order. The feature exemplars that appear in the first position (i.e. the search routine rated them as the closet match to the description) are displayed together as the 'initial face'. If a witness was unable to provide a description of a particular feature the description box is left empty. Any description boxes which are left empty prompt the program to select a default exemplar which is considered as "average". Once this process is completed for all of the facial components, the initial composite face is displayed.

Stage 5 - Modifying an Initial Face

Once an initial face is displayed, the witness uses recognition skills to suggest any amendments to make it resemble their memory of the suspect as closely as possible. A number of options are available. The first stage of modification usually involves a witness selecting a feature and viewing alternative exemplars of it. The operator will scroll through exemplars of a feature and any number of exemplars the witness thinks resembles that of the suspect can be 'tagged'. This aims to reduce the number of exemplars that need to be viewed before finding the best likeness. The operator then scrolls through the subset of tagged features to select the best match to the description of the suspect. Each feature exemplar is always shown within the context of a whole face, never in isolation. Although there is no prescribed order for assembling the parts of the face, Turner (2004) provided some evidence that simply constraining the order of feature selection was beneficial and suggested that working on higher salience features (as described in Section 3.4.1) early in the construction process might lead to the highest quality likenesses.

A witness may move a selected feature up, down, left or right. Paired features (Eyes, Eyebrows and Ears) or groups of features (e.g. Nose with Mouth, Eyes with Eyebrows and so on) can also be moved closer together or further apart. Features may also be resized (longer, shorter, wider or narrower) and re-toned (darker or lighter). The moving, re-sizing and re-toning of features affects all of the exemplars of the altered feature and so any exemplars viewed after these changes have been made will have the same change applied. Therefore, these changes are usually done after the witness has selected the feature which most resembles that of the suspect. Finally, more elaborate

changes can be made to an image with the use of graphics software (e.g. Micrografx Picture Publisher)²⁷.

Stage 6 - Final Facial Image

After the witness has directed all of the modifications they wish to make, a final facial image is displayed, which is the one the witness believes is as good a likeness to the suspect as possible, to enable someone to identify the suspect.

4.2.2 THE COMMUNICATION PROCESS BETWEEN A WITNESS AND A COMPOSITE OPERATOR

The use of the E-FIT feature index described in Chapter 3, and the procedure used with E-FIT by operators described above, highlight the importance of the communication process between a composite operator and a witness. Figure 4.2 illustrates a simplified version of this process.

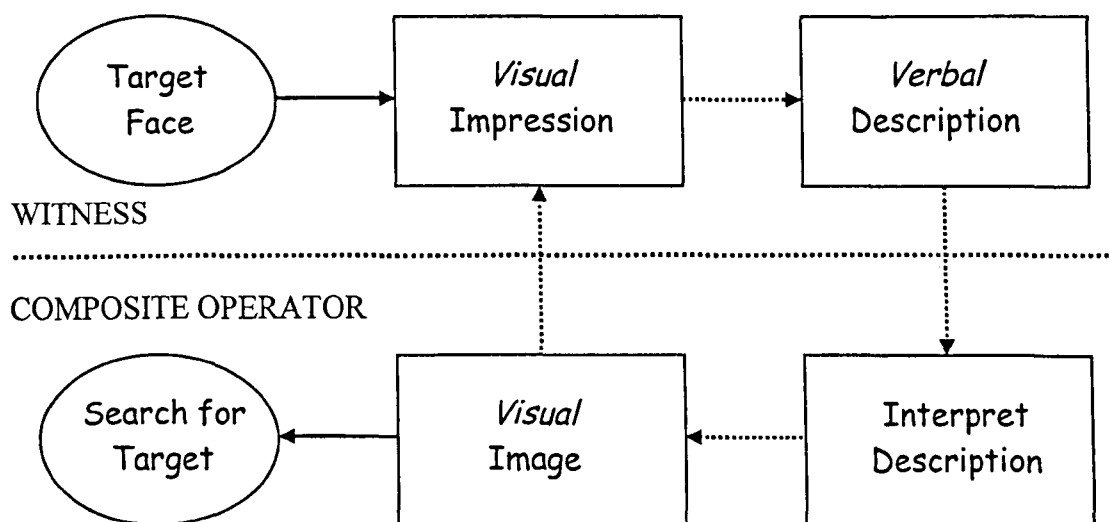


Figure 4.2: The communication process between a witness and an operator²⁸

²⁷ A witness may choose to use none, any or all of the available options to them when altering the initial face.

²⁸ Based on a diagram from Sporer, 1996.

As Figure 4.2 shows, when a witness is attempting to create a likeness there are several cognitive steps involved which form a 'visual verbal' loop which can be described as follows: after a witness has viewed a target's (suspect's) face, they must first recall the visual impression from which they are attempting to create a likeness. They must then translate that visual impression into a verbal description, which must be communicated to a composite operator. The operator has to transform this verbal description back into a visual image. The witness is then presented with an initial visual likeness, and must compare their perception of the likeness to their memory for the target face. From this comparison they must identify and verbalise what changes should be made to the likeness in order to more closely resemble the face they are trying to re-create. This, of course, involves further description and communication stages, followed again by perception and comparison, in a re-iterative process which continues until the witness is satisfied that the likeness is as accurate a representation of their memory for the perpetrator's face as is possible.

This whole process may pose problems at both the level of the witness and the level of the operator:

1) At the level of the witness:

The difficulty with verbal descriptions is that a witness may lack the vocabulary to describe a face. As described in Sections 2.1.3 and 2.1.4, this issue is especially true for child witnesses. Children may also have problems with their comprehension of language: can children understand the questions and prompts being provided by an operator? Previously for this reason it has been suggested by Sporer (1996) that pictures may be used to illustrate descriptions: "providing witnesses with a visual display of a range of alternatives" (p. 57) to avoid linguistic deficits and lead to more accurate accounts. In particular, exploring new techniques such as adjective check lists, rating

scales or alternative choice visual displays might prove useful in making this communication process between a witness and an interviewer more effective. As discussed in Section 2.3.3 props and prompts are potentially useful in helping children to communicate information (Pipe *et al.*, 2002).

2) At the level of the operator:

In terms of children's language comprehension, the main issue for operators concerns the information available to them to question and prompt a child witness. Research introduced in Section 2.1.3 demonstrated that adults not only ask questions that are developmentally inappropriate, but they may also misinterpret children's responses. In terms of children's language production, the main issue for an operator is how accurately they can interpret and translate a child's description of an unfamiliar face based on their own interpretations, to match the E-FIT Index (A. Parry and C. Charsley, personal communication, 2nd November 2001). A study by Allen, Towell, Pike and Kemp (1998) found that E-FITs constructed by operators were significantly better likeness than E-FITs created by operators with a describer. Allen *et al.*, stated that their results suggest that the quality of the E-FIT images is limited by the translation of the verbal description provided by the witness. There are a number of problems in relation to the need for translation of descriptions, which were introduced in Section 3.4.2. To recap: first, the descriptions were devised by adults for adults - there are no comparable descriptions for children; secondly, witnesses do not provide 'standard' descriptions i.e. what one person describes as long hair, may be described as short by another, which may be more likely among child witnesses.

4.2.3 TRAINING OF THE COMPOSITE OPERATOR

In their chapter on face recall, Shepherd and Ellis (1996) stated that "effective training of interviewers may improve the quality of information obtained from witnesses"

(p.112). Indeed, composite operators are mentioned in ABE. The section 'Planning Interviews with Children' includes the guidance states that "in some circumstances, the child witness may be required to perform an identification, or to collaborate with police artists, or facial composite operators.....Police officers carrying out such procedures with child witnesses should be aware of the guidance contained in this document, and may require additional training or support" (Home Office, 2002, p.10).

There are a number of courses available to police operators providing specific training for constructing facial composites. These courses may be received via police forces, via Aspley Ltd. or via the National Training Centre. The main focus of existing facial composite training courses is on interviewing *adults* in order to construct a composite image. There are also a number of courses available which provide specific training for conducting interviews (as described in Section 2.3.4). These courses cover information about interviewing *children* as well as adults. Clark (2000, 2002) presented evidence that not all operators were rigidly adhering to the training in their police work. For example, operators reported showing witnesses E-FIT's description boxes in place of a CI. On a practical point, police operators vary greatly in their background and experience. Some operators are constructing composites full time, whereas, for other operators constructing composites is only part of their duties.

At the time of writing there is no single, standardised training program for composite operators and no formal training course available which provides guidance on interviewing children in order to construct a facial composite. Operators are required to draw on their past experiences when interviewing a child witness. However, some operators may not have much experience, even with adult witnesses. An important step forward would be the provision of a training course in this area. As Taylor (2000)

suggested “every effort should be made to standardise procedures within loosely structured guidelines. This will help ensure composite images will be acceptable as evidence for use in the prosecutorial process in court” (p.164).

4.3 COMPOSITE SYSTEMS RESEARCH

4.3.1 THE EVALUATION OF COMPOSITES

When evaluating composites there is an important distinction between the subjective *likeness* and the objective *utility* of a composite. A composite could be an extremely accurate likeness of a target, however, it may not be useful to the investigation if it is not recognised. Alternatively, a composite may be a fairly inaccurate likeness but may still bear enough of a resemblance to the suspect to be useful. Subjective likeness methods of measuring the accuracy of composite include asking independent judges to rank a set of composite's in relation to the original target face from which they were constructed (a 'ranking' task). Judges may also be asked to assign a score on a scale indicating how much they think that the composite looks like the target photograph (a 'rating' task). Objective measures of a composites utility include asking judges to assign a set of composites to a set of original target faces from which they were constructed (a 'sorting' task) or asking judges to select a target from which a composite has been constructed from among a set of distractor faces (an 'identification' or 'matching' task). If judges familiar to the targets are used they may be asked to provide names for any composite they recognised (a 'naming' task).

Results from research studies using such measures of quality of composites have typically found recognition rates to be poor, and generally not much better than chance. An early study by Ellis *et al.* (1975) examined adult participants' abilities to construct a facial composite using Photo-FIT. Judges' rated the likenesses of the composites to the

target on a scale of one to seven. Judges overall accuracy was just 12.5% which increased to 25% if second and third choices were taken into account. However, as described in the Section 4.1.3, research generally shows serious limitations in the effectiveness of non-computerised composite systems in producing accurate representations of faces.

In a more recent study by Koehn and Fisher (1997), participants observed a target and after a two day delay either verbally described or constructed a facial composite using the computerised composite system Mac-a-Mug from memory. The composites were rated for likeness on a scale of one (not a good likeness) to ten (good likeness). The utility of the composites was also measured by asking judges to identify the target in a target present or target absent photo array. Again, results showed that the rating and the utility of all of the resulting composites was extremely low: 57% of the composites were rated as one and 12% were rated as 2; in the identification task, only 7% of the target present participants made the correct response (when chance was 17%). Kovera *et al.* (1997) also confirmed the low evaluation of composites created from memory using the Mac-a-Mug system. Although this is a computerised composite system witnesses are still required to work on the basis of the selection of individual features in isolation. As described in Section 4.1.3 Koehn and Fisher (1997) speculated whether a composite system that encouraged a more holistic approach to face construction (e.g. E-FIT) might yield higher quality images.

The limited research which has been conducted with the evaluation of composites using the E-FIT composite system has produced mixed results. As described, Davies *et al.*, (2000) only found a significant advantage of E-FIT over Photo-FIT when a participant constructed a composite with the target photograph present. E-FITs constructed from

memory of an unfamiliar face were correctly matched to target on average by 7.5 out of a maximum of 12 independent judges, compared to an average of 7 judges for composites constructed using Photo-FIT. More recently, Turner (2004) conducted a series of laboratory-based experimental studies of E-FIT with adult participants and showed slightly more promising results of the quality of composites: in a matching task, between 29% and 43% of participants correctly matched composites to target.

In comparison to the number of studies conducted with adults, there are far fewer studies which have examined children's abilities to give descriptive statements or construct facial composites. Composite research with children will be considered next.

4.3.2 COMPOSITE SYSTEMS RESEARCH WITH CHILDREN

"There seems to be little agreement among police officers regarding the competence of children as witnesses to facial appearance" (Flin, Markham and Davies, 1989, p.132). Yet Davies *et al.* (1983) reported that there is a belief among police that children may show a particular proficiency at producing facial composites. A New York police artist is reported as saying "kids make the best witnesses. They tend to notice things and aren't afraid to say 'he has a big nose.'" (MacDonald, quoted by Flin *et al.*, 1989). As mentioned in Chapter 3, few studies have explored children's facial recall abilities and even fewer studies have investigated children's abilities to construct facial composites.

Flin *et al.* (1989) conducted a laboratory study of children aged 8- to 9-years and 11- to 12-years to examine the ability of children to construct a Photo-FIT under two conditions. In the first condition, children were presented with a photograph of an adult male face. With the photograph remaining present, they were asked to provide a verbal description, followed by the construction of a composite using Photo-FIT. In the second

condition the same children were presented with a photograph of another adult male face for 60 seconds. The same children were again asked to provide a verbal description, to construct a composite, this time from memory. Finally, children were asked to make an identification of the two targets from a photographic array of 24 faces. The quality of the composites was measured by the ability of judges to sort the composites into piles corresponding to the targets. The performance of children was compared to the performance of 24 adults in a study by Christie, Davies, Shepherd and Ellis (1981).

For both the composite construction and the verbal description tasks there was a consistent and significant improvement with age with competence levels for all ages above chance. On average, adults' Photo-FITs were matched to the correct target 50% of the time, this figure was 40% for the 11-year olds, and 34% for the 8-year-olds (when chance was 17%). This difference between adults and the children was significant and the difference between the 11-year-olds and the 8-year-olds composites approached significance. The overall level of competence displayed by participants of all ages at the face construction task was low.

Results showed no significant difference between the photo absent and the photo present condition for all age groups. As described previously, such a finding has been reported for adults with both Photo-FIT (Ellis *et al.*, 1978) and Identikit (Laughery and Fowler, 1980), and appears to reflect the inflexibility of composite systems rather than memory problems of the witness.

Judges were also asked to sort the verbal descriptions from the face absent description into piles corresponding to the target. Again, results showed a significant effect of age

and interestingly, results showed that the initial verbal descriptions provided a better guide to likeness than the composites for *all* age groups. This result has also been reported for an adult sample (Christie *et al.*, 1981). On the basis of the research described in Chapters 2 and 3 concerning children's language abilities, and in particular their restricted vocabulary in facial descriptions, it is surprising to note that descriptions were superior to composites even among the youngest children. In addition to these results, the accuracy of the description did not correlate with the quality of composite for either the child or the adult comparison sample. Such findings suggest that the quality of composite may well be limited by the difficulty in translating information contained in the verbal description into an accurate visual image (Brace, Pike and Kemp, 2000). All of these results provide further evidence for the insensitivity of the composite methods of facial recall.

Overall, children of 8-years and above were shown to be capable of constructing face composites from the Photo-FIT kit. Flin *et al.* (1989) stressed that the Photo-FITs produced by the youngest group were still sorted at an overall standard which was markedly above chance and that among these composites were a number of excellent likenesses.

In the task used in the study by Flin *et al.* (1989) participants were required to construct a composite immediately following the observation of a photograph, rather than a naturalistic event. A study by Davies *et al.* (1989) investigated whether children would perform better in a more forensically realistic setting where there was a live interaction with an adult. Children aged 6- to 7-years and 10- to 11-year-olds took part individually in a simulated health check procedure. One week later children were asked to provide a free description followed by a prompted description concerning the event and the

appearance of the target. Children were prompted using direct questions relating to the five major aspects of facial appearance represented in Photo-FIT (Hair, Face, Eyes, Nose and Mouth). Half of the children then constructed a facial composite using Photo-FIT. The quality of the composites was rated on a scale of one (poor likeness) to five (very good likeness) by twelve judges. The results of the study showed that there was no significant effect of children's age on composite quality, and judged likeness was generally poor for both age groups: the average score for 6- to 7-year-olds was 1.53 (30.5%) and for 0- to 11-year-olds was 1.67 (33.4%). Davies (1996) reported that these results are in contrast to the results found by Flin *et al.* (1989), however, Flin *et al.* only found a significant difference between adults and children's Photo-FITs, not between children of different ages. This cannot be explored in the study by Davies *et al.* (1989) due to the absence of an adult control group.

Davies *et al.* explain the poor likeness ratings given to all composites by stating that children may simply have failed to observe the individual concerned. However, as Davies *et al.* themselves point out, such an explanation does not account for errors of commission which occurred. They concluded that the possible areas in which children may be lacking most likely lie in either the retrieval of specific facial details or in the translation of such details into verbal terms, or both.

A more naturalistic study was carried out by Schwartz-Kenney, Norton, Chalkey, Jewett and Davis (1996). They examined the ability of children aged 5- to 6- and 8- to 9-years to construct a facial composite using the Identikit system. Children played with a female research assistant for 15 minutes. This interaction was followed immediately by an interview consisting of direct questions concerning the target's age, height, weight and hair colour. Children then constructed a composite, and later 74 judges rated the

composites on a scale of one (very much unlike) to six (very much alike). Although this study did not include an adult comparison group, in contrast to the above studies, the results showed a significant effect of age: older children's composites were rated as significantly more accurate than younger children's composites. The reason for this significant finding could be explained by the inclusion of the 5- to 6-year-olds age group, which is younger than the age groups included in the previous studies. As with the previous research with both children and adults, the likeness ratings were still low with older children's composites rated on average as 2.57 (42%) and younger children's as 2.12 (35%).

In each of the three studies conducted with children described above, it is important to note that the overall mean ratings of likeness were low for all the children, regardless of age. However, the studies conducted by Davies *et al.* and Schwartz-Kenney, Norton *et al.* did not include adult comparison groups and it is necessary to compare children's performance to that of adults. Davies (1996) states "in general, adult competence at composing a facial likeness is not high... and children's performance needs to be judged in that light" (p.243-244). Although the overall likenesses of composites were low in the studies described above, the only measured of accuracy of the composites was a subjective measure, requiring judges to rate the likeness of the composites to the target. Neither of the studies examined the more realistic and objective measure of utility of composites, by asking judges to identify or match the composites to the correct target.

As illustrated above, there have only been a limited number of studies of children's abilities to construct facial composite images. Therefore, the ability of children to use composite systems is still in question. Studies have not found a significant difference

between children of different ages, and the only study which included an adult comparison group (Flin *et al.*, 1989) showed that children were less effective than adults at making composites. Although, children of all ages were capable of making composites and Davies (1996) provides a useful summary of this research and concludes that “children are capable of making facial likenesses but their general competence appears lower on average than adults” (p. 245). Importantly, where an adult comparison group was included, children’s composites were included amongst the best likenesses rated by judges. In other words, these results suggest that adults do not consistently produce facial composites of a better quality than children. Large variations within all age groups results in some composites constructed by children of better quality than some composites constructed by adults.

4.3.3 IDENTIFICATION AFTER CONSTRUCTING A COMPOSITE

After a witness has constructed a facial composite, they may be called upon to identify a suspect in line-up or through photographs. As Section 3.4.3 described, some research has found that the verbalisation of a facial description may interfere with later identification performance (e.g. Schooler and Engstler-Schooler, 1990). It is therefore necessary to briefly consider previous research which has investigated the effects of composite production of subsequent identification performance of an eyewitness.

Early research illustrated negative effects on the identification performance of witnesses who had previously worked with sketch artists to construct composites (e.g. Brown, Deffenbacher and Sturgill, 1977). However, studies which have examined the effects of constructing a composite on a subsequent recognition task using the Identikit systems have produced contrary findings. Mauldin and Laughery (1981) found that participants who constructed Identikit composites of the target face were more likely to select the

target face than participants who did not form composites in a later recognition task. However, this study can be criticised for not using unbiased line-up instructions (i.e. telling the participants that “the target person may or may not be present”) and all of the participants were required to choose an alternative from the photo array. In a more forensically realistic study by Yu and Geiselman (1993) participants viewed a videotape of a simulated robbery and either constructed an Identikit composite, provided a written description or were in a control condition. After a two day delay, participants were presented with both target-present and target-absent photo arrays. Yu and Geiselman found that building an Identikit composite did not interfere with later identification performance from a photo array. Pike and Brace (2002) also reported similar results with the E-FIT system. In a laboratory experiment 73 adult participants saw a video of a staged crime scenario. 37 of these witnesses constructed a composite with the researcher. A field experiment was also conducted where 42 adult witnesses viewed a staged crime in a small car park. Half of these participants went on to produce a composite image with a trained E-FIT operator. The results from both the laboratory experiment and the field study suggest that creating a composite does not impair performance on a later identification.

The results of the studies cited above with the Identikit and E-FIT systems may be used to disregard any cautionary note concerning eyewitness identification performance following the production of a composite with a sketch artist (Brown *et al.*, 1977). Yu and Geiselman proposed that “the activity of forming composites with mechanical systems may be qualitatively different in its effects on memory for a face than working with a sketch artist” (p.289).

As with verbal overshadowing research (section 3.4.3), the majority of studies in this area have been conducted with adults. The study by Davies *et al.* (1989) described in section 4.3.2 went on to examine children's subsequent identification performance of a target from an array of eight faces. For half of the children this identification was made after a verbal description and for the remaining half the identification was made after a verbal description and the construction of a photo-FIT face. Results showed that the construction of a Photo-FIT had no effect on the subsequent identification task.

In summary, the limited research which has been conducted shows that the construction of a facial composite using a mechanical (non-computerised) system does not have a detrimental effect on subsequent identification performance (Davies *et al.*, 1989; Yu and Geiselman, 1993) and may actually facilitate it (Mauldin and Laughery, 1981) (when witnesses are not provided with a 'not present' option in the line-up). It remains to be seen what effect constructing a composite with a computerised system may have on subsequent identification performance for both adults and children²⁹.

4.4 CHAPTER 4 SUMMARY

This chapter considered the practical issues surrounding the procedure of obtaining facial composites from child witnesses. The communication process between a witness and an operator was described in more detail than in the previous chapters and it was argued that appropriate interview techniques may assist both operators and witnesses in the production of facial descriptions and subsequent composite production.

²⁹ An identification task was included for Study 3 and 4 of the thesis in order to determine whether or not child participants had attended to the target.

In comparison to the number of studies conducted with adults, there are far fewer studies which have examined children's abilities to construct facial composites. Although the limited research with non-computerised systems demonstrated that children's general competence at constructing facial composites is lower than that of adults, it was argued that children are able to construct facial composites. The studies cited found no statistically significant difference between children aged between 5- and 12-years. In the one study where an adult comparison group was included there was a statistically significant difference between adults and children. Although importantly, some children's composites were judged to be among the 'best' likenesses produced, suggesting that adults do not consistently produce facial composites of a better quality than children.

Further research is needed which investigates the newer, computerised composite systems such as E-FIT, which is more compatible with the holistic processing that is said to characterise facial processing for adults and more recently for children (as described in Chapter 3). As such, one of the central aims of this thesis was to examine children's abilities to construct composites with the computerised system E-FIT.

4.5 THESIS RESEARCH QUESTIONS

In the preceding chapters of this thesis the communication process between an interviewer/operator and a child witness was focused on. It was argued that a detailed consideration of the language and terms children use to describe unfamiliar faces may assist this communication process. It was also argued that an appropriate interview technique may be used to support both operators and witnesses in the production of facial descriptions and subsequent composite production.

The focus of the research presented in the following chapters describes experimental work undertaken in order to: gain an understanding of children's verbal descriptions of unfamiliar faces; and, to explore how children and operators might be assisted through appropriate interview techniques to provide computerised facial composite constructions of unfamiliar faces. Specific research questions which emerged from the literature review are listed below along with the relevant study or studies which sought to address them.

(1) *How effectively can children produce verbal descriptions of an unfamiliar face?*

The term *effectively* in the research question refers to both the content and quantity of descriptions useful for subsequent composite construction. In the present research, Study 2 was designed to investigate the content (i.e. what words children use to describe faces) and quantity of children's verbal descriptions of unfamiliar faces.

(2) *Is there an appropriate interview technique which can be used to enhance children's facial descriptions for subsequent computerised facial composite constructions?* The thesis involved the development of a set of visual and verbal prompts as potentially appropriate interview techniques to assist witnesses and operators in obtaining facial descriptions for subsequent composite construction. Study 3 investigated the effect of these prompts on the content and quantity of children's verbal descriptions of unfamiliar faces.

(3) *Can children produce computerised facial composites of an unfamiliar face when using an appropriate technique to enhance their facial descriptions?* Study 4 went on to

explore the use of the prompts described above on children's abilities to construct facial composites using the computerised system E-FIT.

In the following chapters the theory and research relevant to each of the studies is summarised. Each study outlines specific research aims including some additional aims. Throughout the thesis the effect of different age groups on each of the aims was explored. A first step in the research was to establish composite operators' current practice with child witnesses. Study 1 was designed to address this and is described in the following chapter.

CHAPTER 5: STUDY 1

A QUESTIONNAIRE SURVEY OF COMPOSITE OPERATORS

This chapter describes a questionnaire survey of E-FIT operators to identify current practice with child witnesses and examine any practical issues faced.

5.1 INTRODUCTION

As described in Chapter 4, at the time of writing there is no single training course available to composite operators which provides guidance on interviewing children in order to construct a facial composite. The absence of such a training course, along with the lack of research examining children's abilities to provide verbal descriptions and construct facial composites, implies that there is no 'typical' interview structure for producing composites with children. Consequently, there is a need to establish existing procedures currently used by composite operators with child witnesses. Therefore, the main aims of Study 1 were to:

- (i) identify operators' current practice in creating composites with child witnesses;*
- (ii) elicit operators' experiences with, and opinions of, child witnesses;*
- (iii) identify any practical issues faced by operators when obtaining verbal descriptions and computerised facial composites from child witnesses.*

All of these aims were treated as exploratory and no hypotheses were stated.

5.2 METHOD

5.2.1 DESIGN

Construction of the Questionnaire-Based Survey

In order to achieve the aims a questionnaire-survey was designed, which was informed by a number of sources. First, a series of informal interviews were conducted with two Metropolitan police composite operators who have had extensive experience of interviewing children in order to construct facial composites. Operators were asked about their current practices with adult and child witnesses and the practical issues they felt they faced when interviewing children including: training, current interview procedure, offences they interview for, 'typical' suspects, use of prompts and issues with subsequent identification. Secondly, the author was a participant in a composite interview with an experienced composite operator. Finally, the author observed a demonstration of a facial composite interview conducted by an experienced operator with a 10-year-old girl.

The interviews, reflections on the personal experience of being interviewed, and the demonstration interview, raised a number of questions concerning current police practice and experience with child witnesses which highlighted the importance of drawing on a 'practical perspective'. Thus the questionnaire was constructed based upon both the practical issues which arose from the discussions with police operators and the theoretical issues described in Chapters 1 through 4.

Pilot

A questionnaire was sent to four people with experience of E-FIT. These were a police officer, a research student who was investigating the use of E-FIT with adults, the Head of Aspley Ltd. (who supply E-FIT) and the Head of the Association of Chief Police

Officers (ACPO). These participants were asked to comment on the construction of the questionnaire, complete the questionnaire and to provide any notes which would assist future participants in the completion of the questionnaire. On the basis of this information some minor revisions were made to the questionnaire.

5.2.2 PARTICIPANTS

The questionnaire was sent to 277 E-FIT operators who were registered with Aspley Ltd. in January 2002³⁰. Fifty-three operators (37 males (70%), and 16 females (30%)) completed questionnaires. Operators ranged in age, from 30- to 59-years, and in rank, from Constable to Detective Inspector.

5.2.3 MATERIALS

The final questionnaire consisted of an introductory page explaining the reasons for the survey and introducing the author. This was followed by seven sections. Each section included a variety of question styles, namely open, closed, 'check all that apply', and 'rank on a scale' type questions. The section names and an indication of the topics covered in each section are summarised (a full copy of questionnaire can be viewed in Appendix II):

<i>Section</i>	<i>Details of Questions</i>
<i>1 Personal Details</i>	Asked for personal details of the operator – age, sex, length of time they had been a composite operator, their experience with children (both as part of their job with the police and also in general).
<i>2 Children and Composites</i>	Covered information about the number of composites operators construct with children. Operators were also asked the ages of children they had interviewed as well as the age of child they believe it is possible to interview.

³⁰ This number included *all* police who registered with Aspley Ltd. and was therefore an overestimate of the number of operators actually constructing composites with witnesses.

<i>3 Training</i>	Asked operators to detail any training (both formal and informal) they had received for interviewing both adults and children in order to construct composites. Operators were also asked to report any other interviewing training they had received.
<i>4 Time</i>	Required operators to estimate the time they spend with child witnesses and to compare this to the time they spend with adult witnesses. Operators were also asked to report whether they thought the time they spend was appropriate for adult and child witnesses.
<i>5 Interview Stage</i>	Asked operators questions about their procedure in the interview stage - the type of questioning they use, any problems which occur in this stage with children and any additional materials operators use.
<i>6 Construction Stage</i>	Invited operators to report their procedure in the composite construction stage - the techniques they use on the computer and any problems which occur in this stage with child witnesses.
<i>7 Opinions about Child Witnesses</i>	Asked for operators' experiences and opinions of child witnesses in general. This section included questions about features children appeared to be good or poor at recalling, any changes they would like to see made to whole procedure, and to provide an outline of their preferred method of producing composites with children.

5.2.4 PROCEDURE

In January 2002, questionnaires were sent out via Aspley Ltd. to ensure operators' personal details remained anonymous. Operators were asked to return completed questionnaires in the return postage paid envelope provided within three months. Personal details of the author were given to operators in case of comments or queries. Findings from the questionnaire were disseminated to operators through conference

presentations (Paine *et al.*, 2002; Paine *et al.*, 2003b; 2003c) and a workshop (Paine, *et al.*, 2004).

5.3 RESULTS

The results section is organised by the questions asked in each of the seven main sections of the questionnaire survey. The number of operators who responded to each question (written as n) is indicated throughout the section.

5.3.1 TREATMENT OF THE RAW DATA (CODING OF RESPONSES)

After the final date for the return of the questionnaires, all responses were collated (n=53) and entered into a statistical analysis package (SPSS). Responses to closed, 'check all that apply' and 'rank on a scale' questions were quantified. Detailed responses for open questions were considered before a smaller set of categories were created into which each responses could be sorted (i.e. content analysis on open question responses). For a number of questions operators were able to list more than one category or reason. Therefore, totals for these questions do not always equal 100%.

5.3.2 SECTION 1: PERSONAL DETAILS

Which composite package are operators currently using?

(n=53) All respondents reported that they were currently using the computerised facial composite system E-FIT (version 3.1a) to construct facial composite images with witnesses.

How long have respondents been facial composite operators?

(n=53) The data indicated that operators varied greatly in their background and experience. Responses ranged from 6 months to 204 months (17 years); the mean length

of time respondents had been an operator was 67.7 months (just over 5 and a half years) and mode = 36 months (3 years).

What experience have operators had interacting with children?

Operators were asked about their experience interacting with children first, as part of their job with the police, and secondly, in general. Responses indicated that few operators (11.8%) had daily contact with children as part of their job ($n=51$). However, in terms of operators' experience of interacting with children in general, the majority of operators (81.6%) stated that they were parents ($n=49$).

5.3.3 SECTION 2: CHILDREN AND COMPOSITES

On average, how many composites do operators construct in a year?

($n=53$) There was a wide range of responses to this question with operators constructing between 3 and 150 composites a year. For some operators, the construction of E-FITs was a full time job, whereas for other operators this was only part of their duties. The mean number of composites produced was 25.6, with the majority of operators constructing around 20 composites a year.

On average, how many composites do operators construct in a year with children?

($n=53$) As with the previous question, there was a wide range of responses. Some operators reported only constructing 1 composite a year with a child, whilst other operators stated they were constructing up to 18 composites a year with children. The mean number of composites operators constructed in a year with children was approximately 3 and the majority of operators were constructing 2 composites a year with children.

The number of composites an operator constructed with child witnesses in a year was calculated as a percentage of the total number of composites they constructed in a year. On average, 18% of composites were constructed with child witnesses. Again, there was a wide range in responses from 3% to 67% indicating that, for some operators the majority of the composites they constructed in a year were with children, whereas for other operators very few composites were constructed with children. For the majority of operators, 10% of the composites they constructed were with children.

Why do some operators not construct composites with children?

(*n*=19) Seventy-two operators returned completed questionnaires although only 53 of these operators had constructed composites with children. Of the 19 operators who reported that they do not construct any composites with children, all but 1 operator stated that the reason for this was that “the need has never arisen” or that a composite interview with a child “has never been requested”. These reasons were provided even though the mean length of time these operators had been constructing composites with adults was just over 3 years and 1 respondent had been constructing composites for 9 years.

Further reasons for not conducting E-FIT interviews with children included a lack of facilities; for example, that “interview facilities are not appropriate for very young children” and that “E-FIT in general has not been encouraged due to a lack of operators and available equipment”. However, other reasons provided were simply statements about child witnesses which were not supported by any evidence. For example, “[I] have real doubts about the ability of young children to recall faces or describe them so would normally look for other options”; “officers do not trust the judgement of children to request an E-FIT”; “I was taught that children below 8-years of age would not be very good for E-FIT”.

Who or what determines at what age a child can construct a composite?

(*n*=52) The majority of operators (59.6%) stated it was their decision at what age a child can construct a composite, 17.3% stated it was a superior officer's decision and 23.1% stated it was a joint decision.

What is this decision based on?

Operators were asked to rank on a scale of 1-5 what this decision is based on, in terms of the abilities of a child. Table 5.1 shows the percentage of operators who rated each ability at each level of importance on their decision. A score of 1 indicated low importance and a score of 5 indicated high importance.

		% of operators				
		Importance				
Decision is based on ability of child to:	<i>n</i>	<i>low</i> 1	2	3	4	<i>high</i> 5
Remember a face accurately	52	0	1.9	21.2	34.6	42.3
Describe a face accurately	52	0	3.9	19.2	32.7	44.2
Understand what is required of them	50	0	6	14	38	42
Be questioned fairly	50	4	6	36	34	20

Table 5.1: Factors operators reported influence their decision to ask a child to construct composite in response to Study 1

From Table 5.1 it can be seen that operators seemed to agree on their rating of children's ability to *remember a face accurately*, to *describe a face accurately*, and to *understand what is required of them* can be seen. The majority of operators rated all of these as of fairly high to high importance (a score of 4 or 5) when deciding whether children should be approached to construct a composite. However, the importance of the ability of the child to *be questioned fairly* was rated differently: a lower percentage of operators rated this of high importance, and some rated this of low importance. Interestingly, this is the only reason which included the abilities of the operator (by mentioning questioning).

In addition, 16 operators stated 'other' reasons of high importance (from 4 to 5). These reasons included the confidence of the child and the concentration/attention span of the child.

What is the age of the youngest child operators have interviewed in order to construct a composite?

($n=53$). Figure 5.1 illustrates the percentage of operators who have interviewed a child of each age shown.



Figure 5.1: Percentage of operators who reported interviewing a child of each age in response to Study 1 ($n=53$)

Figure 5.1 shows a large variation in the youngest age of child operators have interviewed in order to construct a composite. Around 4% of operators reported that the youngest age of child they have interviewed in order to construct a composite was 6-years. For around 10% of operators, the youngest age of child they reported interviewing was almost 10 years older, at 15-years of age. Less than 20% of operators stated they had interviewed a child under 10-years of age whereas almost 70% had interviewed a child over 10-years of age.

What is the youngest age at which operators think children can construct a composite?

($n=43$) Figure 5.2 shows the percentage of operators who believe it is possible to interview a child of each age shown.

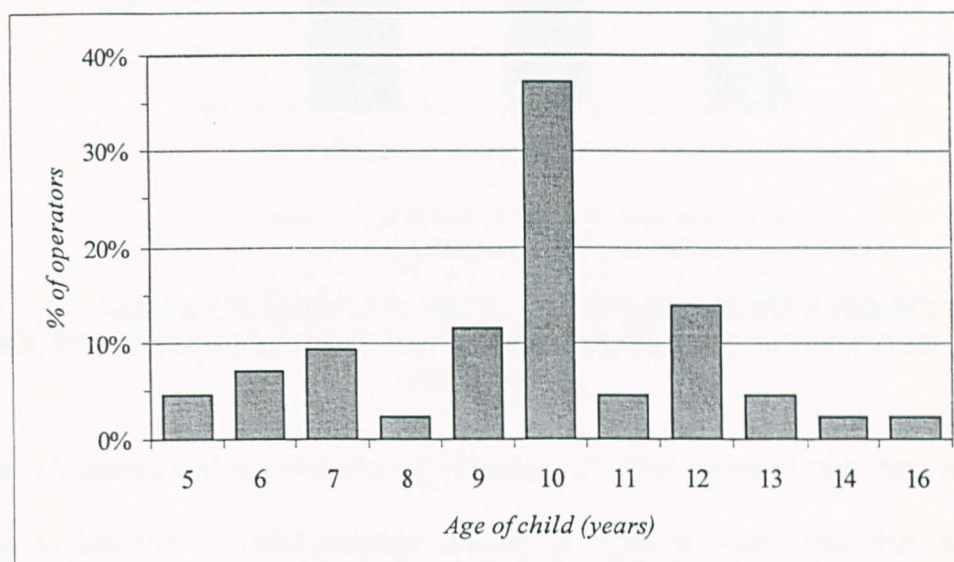


Figure 5.2: Percentage of operators who stated they believe it is possible to interview a child of each age in response to Study 1 ($n=43$)

Figure 5.2 again shows a range of responses, from 5- to 16-years of age. However, 37% of operators stated that 10-years was the youngest age of child they would be willing to interview. Importantly, around 5% of operators believed it was possible to interview a 5-year-old child in order to construct a composite.

A comparison was made between the age of child each operator reported they believe it is possible to interview (Figure 5.2) and the youngest age of child they had interviewed (Figure 5.1). Figure 5.3 shows this comparison.

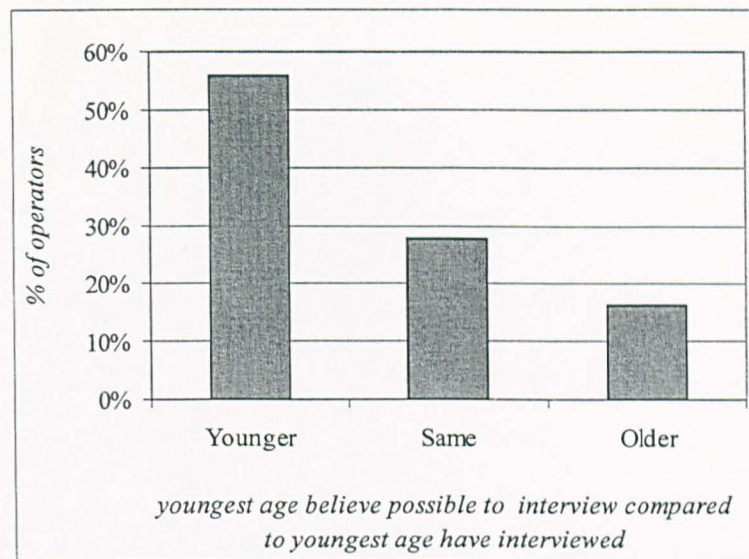


Figure 5.3: Comparison between the age of child operators reported they believe they could interview and the youngest age of child they have interviewed in response to Study 1 (n=43)

Figure 5.3 shows that the majority of operators (55.8%) reported that they would be willing to interview a child younger (mean=3.2 years younger) than they have had experience of interviewing. Sixteen per cent of operators stated an older age of a child than they have had experience of interviews (mean= 2.6 years older). Twenty-eight per cent of operators stated the same age for the age of child they thought it was possible to interview and the youngest age of child they had interviewed. These ages ranged from 6- to 13-years.

A number of Pearson's correlation designs were employed to test whether there was any relationship between an operator's experience and the age of child an operator reported they believe it is possible to interview in order to construct a composite. Results showed: a statistically non-significant correlation between *the youngest age of child an operator has interviewed and the youngest age of child an operator believes it is possible to interview in order to construct a composite* ($r=0.180$, $n=43$, $p=0.247$, two-tailed); a statistically non-significant correlation between *the total number of child composites an operator constructs in a year and the youngest age of child an operator*

believes it is possible to interview in order to construct a composite ($r=0.006$, $n=43$, $p=0.970$, two-tailed); and a statistically non-significant correlation between *an operator's experience interacting with children in general (outside of the police force) and the youngest age of child an operator believes it is possible to interview in order to construct a composite* ($r=0.057$, $n=40$, $p=0.725$, two-tailed). Therefore, there was therefore no consistent relationship between operators' experience with children and the age of child operators are willing to interview.

What types of crimes do children construct composites for?
($n=50$) The most frequently listed crimes were:

- sexual offences (including rape and indecent exposure) (68%)
- robbery/theft (50%)
- abduction/attempted abduction (26%)
- assaults (18%)
- burglary (10%)

When asked whether the crimes children constructed composites for differed from the crimes adult witnesses constructed composites for, 70% stated yes and 30% stated no ($n=43$).

5.3.4 SECTION 3: TRAINING

In Section 3 operators were asked about both *formal* and *informal* training they have received for composite production.

What formal training have operators received for composite production?
($n=53$).

- The majority of operators (69.8%) stated they had received training via Aspley Ltd.

- 41.5% of operators stated they received their training through the National training centre
- 26.4% were self taught
- 17 % received training via their police force
- 17 % through other users
- 7.6% through the Scottish police college

The majority of operators attended these training courses approximately 5-years ago and some operators attended courses over 10-years ago. The usual length of training course reported was around two weeks and only 5.7% of operators reported that they updated their training at regular intervals.

No operator mentioned receiving any training for producing composites with children and 51 operators (almost all of the 53 respondents) specifically stated this in the questionnaire. Additionally, 13 operators stated that they would like to receive training for constructing composites with children.

There are courses operators can attend which provide specific training for conducting interviews with adults and children. These include CI (Geiselman and Fisher, 1992) training or MoGP training (Home Office, 1992). Operators were asked to indicate the training they had received specifically for interviewing adults and children. The majority of operators had received CI training for use with adult witnesses (n=45), and 14 operators reported receiving CI Training specifically for use with children. Only 3 operators reported receiving MoGP training for use with children. Twelve operators reported receiving other interview training for use with adults and 5 reported receiving other training for use with children. These included child protection training,

counselling training and sexual offences interview techniques. Again, many operators received training in these interview courses over 10-years ago and some operators reported they received this training almost 20-years ago.

What informal training have operators have sought for composite production with children?

In addition to formal training, 15 operators stated they had sought informal training. Of the 15 who had undertaken less formal training, this included: reading material (e.g. Home Office reports, police reports, psychology books, self help books ($n=7$); general experience/observation ($n=3$); child development course ($n=1$); watching child development video ($n=1$).

Do operators treat children differently to adults in relation to their formal or informal training?

($n=36$) Operators were asked in what ways, if any, they treated child witnesses differently to adult witnesses in relation to their training or experience. The most frequently listed differences were:

- language (e.g. having to simplify and/or modify language; providing more explanations; changing questions and using less formal wording) (83.3%)
- time spent interviewing (e.g. increasing time spent or being more patient; shortening the interview; importance of breaks) (22.2%)
- other ways of treating children differently to adults (27.8%) including interviewing with appropriate adult present; being aware of a child's needs; treating a child with equal respect; interviewing a child at home /or in a child-centred suite; and letting a child work on the computer).

5.3.5 SECTION 4: TIME

Of the time operators spend with a child witness, how much is spent in the different stages of an interview?

Table 5.2 shows the mean time operators reported spending in each interview stage.

Interview Stage	n	Mean Time Spent (mins)			
		Mean	S.D.	Min	Max
Rapport Building	38	19.57	17.27	5	90
Interview (description)	39	37.95	25.10	10	120
Composite Stage	39	66.23	24.48	15	120
Total	37	121.14	38.87	65	210

Table 5.2: Mean time operators reported spending in each interview stage in response to Study 1

From Table 5.2 it can be seen that there was a wide range in responses for all stages. Overall, the mean average time operators reported spending with child witnesses was just over two hours. Responses ranged from 1 hour and 5 minutes for 1 operator to 3 and a half hours for 3 operators. Additionally, operators reported spending on average just over one and a half times as long in the composite stage than in the interview stage.

Do operators spend the same amount of time in each interview stage with a child as with an adult?

Operators were also asked to compare the average time they reported spending with a child to the average time they spend with an adult in each stage of an interview. Figure 5.4 summarises operators' responses.

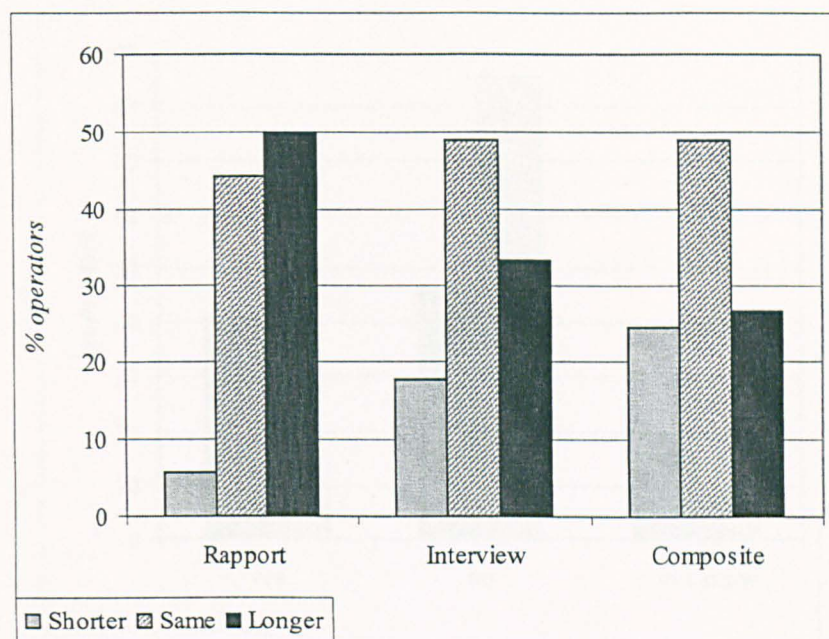


Figure 5.4: Comparison of the time operators reported spending with child and adult witnesses in response to Study 1

Figure 5.4 shows that the majority of operators reported spending the same amount of time with a child witness as with an adult witness during the description interview, and the composite construction stages of the interview. In the rapport stage of the interview, the majority of operators reported spending longer with child witnesses than with adult witnesses. The lowest number of operators reported spending shorter time with child witnesses in this section of the interview.

Do operators think that the whole composite interview process is too long for adults or for children?

Operators were asked if the whole process was too long for adults ($n=49$) or for children ($n=49$). This information is shown in Figure 5.5.

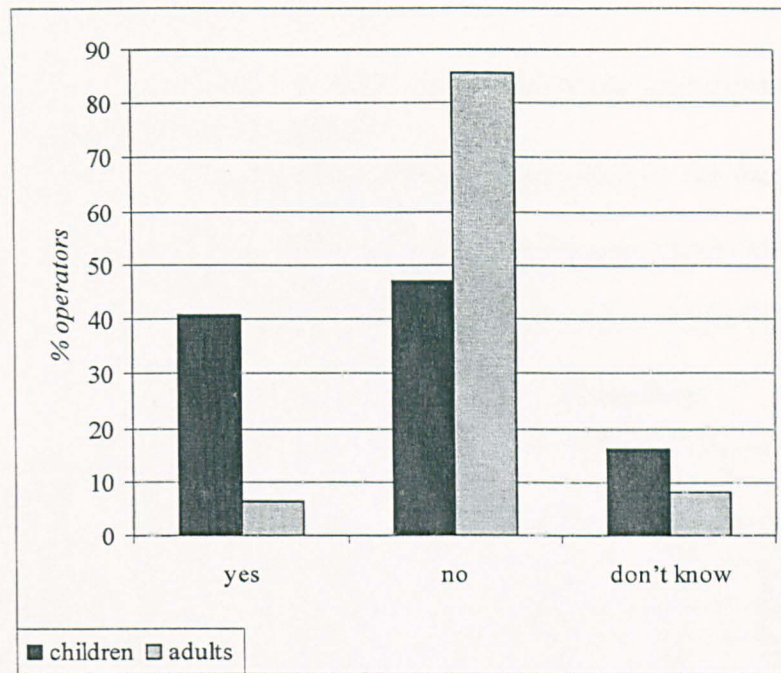


Figure 5.5: Operators responses to the question “Is the composite interview process too long for adults or children?” in Study 1 (n=49)

Figure 5.5 illustrates that the majority of operators felt that the composite interview process was not too long for adults. However, in comparison nearly half (45.1%) of operators did not think that the process was too long for children, this was almost half the number of operators who reported that it was not too long for adults. Many operators (39.2%) reported they thought the process was too long for children.

Overall, those operators who stated that the process was too long for a child, and specified the time they spent with a child witness, spent *more* time, on average, with a child witness (range from 75 minutes to 210 minutes, mean=129 minutes, $n=16$) than those who stated that the process was not too long for use with children (range from 65 minutes to 180 minutes, mean=115 minutes, $n=17$). Furthermore, of those operators who thought the process was too long 7 stated that the whole process should be 1 hour, at maximum. Four operators stated between 1 to 2 hours and 1 operator stated between 2 to 3 hours.

5.3.6 SECTION 5: INTERVIEW STAGE

How frequently do operators use each interview component, and how effective do they think each one is?

For a number of interview components operators were asked to rate the *frequency* (how often they use each one with a child witness) and *effectiveness* (how useful a technique they think it is for a child witness). The findings are shown in Tables 5.3 and 5.4:

Interview component	Frequency (% operators)					
	<i>n</i>	<i>Low</i> <i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>High</i> <i>5</i>
Free Recall	52	5.8	5.8	11.5	25	51.9
Open Questions	51	2	3.9	9.8	23.5	60.8
Specific Questions	51	7.8	11.8	27.5	21.6	31.4
Closed Questions	50	28	18	28	14	12
Questions not relating to face	51	3.9	9.8	27.5	23.5	35.3
Context reinstatement	51	2	3.9	21.6	39.2	33.3
Imagery techniques I	50	0	6	26	30	38
Imagery techniques II	44	4.6	15.9	34.1	25	20.5

Table 5.3: Frequency rating of each interview component provided by operators in response to Study 1

Interview component	Effectiveness (% operators)					
	<i>n</i>	<i>Low</i> <i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>High</i> <i>5</i>
Free Recall	52	3.9	13.5	21.2	42.3	19.2
Open Questions	49	0	14.3	24.5	32.7	28.6
Specific Questions	50	4	8	26	36	26
Closed Questions	47	10.6	12.8	44.7	14.9	17
Questions not relating to face	51	3.9	9.8	37.3	27.5	21.6
Context reinstatement	51	0	9.8	37.3	39.2	13.7
Imagery techniques I ³¹	50	4	8	40	32	16
Imagery techniques II ³²	43	7	20.9	34.9	25.6	11.6

Table 5.4: Effectiveness rating of each interview component provided by operators in response to Study 1

Table 5.3 shows that the interview component reported as being used with the highest *frequency* by the most operators was open questions followed by free recall. Table 5.4 shows the interview component reported as most *effective* by most operators was open

³¹ Refers to operators asking witness to 'picture in minds eye'.

³² Refers to operators asking a witness to 'close their eyes' or to 'look away'.

questions followed by specific questions. Although a high percentage of operators listed that they used free recall with witnesses frequently, the effectiveness of these technique was only rated as high by around 20% of operators.

A number of Pearson's correlation designs were employed to test whether there was any relationship between an operator's report of frequency and how effective they rated a number of interview techniques. Results showed for all interview components operators' reports of frequency were statistically significantly positively correlated with the effectiveness they rated each interview technique³³.

When asked to make comparisons with adults, of 29 operators, 34.5% stated that adults were better in the free recall section of the interview (that their language abilities were better than a child's, therefore enabling adults to provide more detailed free descriptions); and, 17.2% stated that they had to use specific questions more with children (and that children responded well to specific questions). Other differences listed included adults having better concentration levels than children (13.8%) and open questions being used more with adults (10.3%).

What problems occur in the interview stage with children?

98% of operators reported problems ($n=48$). The most frequently reported problems were ($n=47$):

- concentration problems (children being easily distracted or bored) (57.4%)
- language problems (in terms of children describing and understanding) (34.0%)

³³ Free recall [$r=0.595$, $n=52$, $p<0.01$]; open questions [$r=0.399$, $n=49$, $p<0.01$]; specific questions [$r=0.521$, $n=50$, $p<0.01$]; closed questions [$r=0.556$, $n=47$, $p<0.01$]; questions not relating to face [$r=0.783$, $n=51$, $p<0.01$, two-tailed]; context reinstatement [$r=0.335$, $n=51$, $p<0.05$]; imagery techniques I (i.e. picture in mind's eye) [$r=0.433$, $n=50$, $p<0.01$]; imagery techniques II (asking child to close eyes/look away) [$r=0.527$, $n=43$, $p<0.01$], all two-tailed.

- other problems listed included emotional problems (17%) children being too eager to please (12.8%), appropriate adult problems (interruptions or embarrassment in front of) (10.6%) and children exaggerating details (4.3%).

Do operators use any additional materials during an interview with a witness?

(*n*=44) Forty-seven per cent of operators reported using additional materials (e.g. shape charts, size charts, list of descriptors, pen and paper) with adult witnesses and 55% of operators reported using additional materials with child witnesses (including all of the 21 operators who reported using additional materials with adults). Operators from the British Transport Police, West Mercia constabulary and an anonymous operator included the forms they used to collect witnesses descriptions. These forms consisted of one to two word open prompts relating to various facial features (e.g. “Face Shape?”, “Eye Shape?” and so on). Operators from the Metropolitan police included a set of visual prompts they use with both adult and child witnesses. These visual prompts included simple pictures of various shapes and sizes for a number of facial features. These prompts are described in more detail in Section 7.1.

Materials used with adults and child witnesses are described in Table 5.5.

Material	% operators	
	<i>Adults</i> (<i>n</i> =21)	<i>Children</i> (<i>n</i> =24)
Paper and pen	38.1	41.7
E-FIT description boxes	23.8	25
Draw shapes/shape charts	33.3	25
Descriptive form/list of descriptors	23.8	20.8
Draw scale/scale charts	4.8	8.3
Operator - for height/weight etc	4.8	4.2
Operator physically moves	4.8	0

Table 5.5: Additional materials operators reported using in response to Study 1

Table 5.5 shows that operators used various materials with both adult and child witnesses. The most frequently used additional materials with both adult and child witnesses were a pen and paper.

5.3.7 SECTION 6: CONSTRUCTION STAGE

Do operators show E-FIT's description boxes to adult and child witnesses? (n=51) 58.8% of operators showed description boxes to adult witnesses and 41.2% of operators showed description boxes to child witnesses³⁴.

As shown in Table 5.5, approximately one quarter of operators reported using E-FIT's description boxes as an 'additional material'. When explicitly asked whether they used these boxes with witnesses the percentages were much higher. With over half of operators showing the boxes to adult witnesses and just over 40% of operators showing the boxes to child witnesses. The main reasons operators provided for showing the description boxes are provided in Table 5.6.

Reasons for showing witness description boxes	% operators	
	Adults (n=29)	Children (n=20)
To prompt witness for a more detailed description	41.4	45
To focus/involve witness	37.9	45
To clarify witnesses descriptors	17.2	5
Other ³⁵	13.8	15

Table 5.6: Reasons operators stated for showing description boxes to witnesses in response to Study 1

Table 5.6 shows that the reasons given for showing the description boxes were approximately the same for adult and child witnesses. The majority of operators showed the boxes to prompt the witness for a more detailed description. Many operators reported that they used the boxes to focus and involve witnesses.

³⁴ E-FIT's description boxes were described in Section 4.2.1.

³⁵ Included: to show witness the description 'average' can be used; and, 'simply because they are there'

What problems occur in the construction stage with children?

(*n*=37) The majority of operators (86.5%) reported problems with construction, although this was not as many as reported problems in the interview stage (98%).

The most frequently reported problems were (*n*=29):

- *concentration* problems (children being easily distracted or bored) (34.5%)
- *language* problems (in terms of children describing and understanding) (13.8%)
- indecisiveness (children requiring excessive ‘fine tuning’) (13.8%)
- children not treating the system seriously (children playing with the system) (13.8%)
- other problems (44.8%) including: explaining to the child that a composite is not an exact photograph (10.3%) emotional problems (6.9%); children accepting the image too easily (6.9%), feature resizing problems (6.9%); feature moving problems (3.5%); and finally, feature searching problems (3.5%).

The two most frequently reported problems in the *construction* stage with child witnesses were the same two most frequently reported problems in the *interview* stage with child witnesses, namely concentration and language problems.

5.3.8 SECTION 7: OPINIONS ABOUT CHILD WITNESSES

In the final part of the questionnaire, operators were asked general questions about their experiences and opinions of *child* witnesses.

Do operators think that the current procedure they use to obtain descriptions and composites from children is optimal?

(*n*=47) Figure 5.6 illustrates operators’ responses.

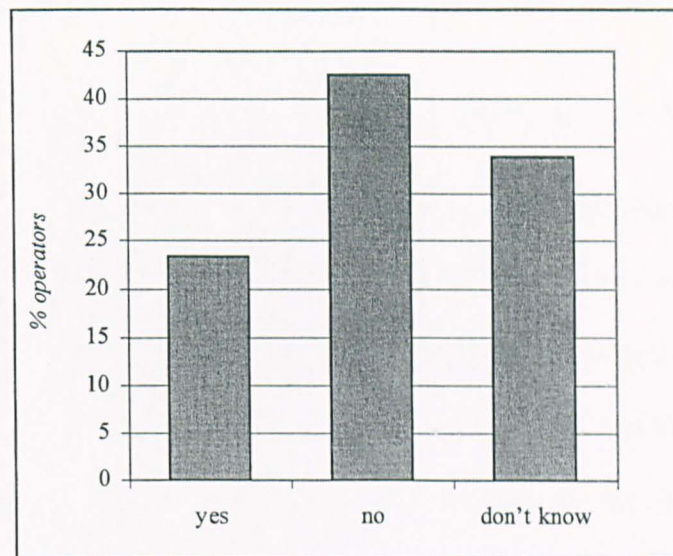


Figure 5.6: Percentage of operators who stated they believe the current procedure they use is optimal in response to Study 1

From Figure 5.6 it can be seen that the majority of operators reported that the current procedure was not optimal. Less than a quarter of operators thought that it was optimal and approximately a third did not know.

What changes do operators think should be made to the current procedure they use to obtain descriptions and composite from children?

(n=27) Changes that operators suggested could or should be made were:

- to receive specific training/a set of guidelines/a set of prompts (51.8%)
- some way to shorten the whole process (14.8%)
- to interview a child without an appropriate adult present (11.1%)
- for changes to be made to the system (i.e. E-FIT) (11.1%) (e.g. advanced shading tools and more overlays such as scars)
- to video interviews with child witnesses³⁶ (11.1%)
- Other suggestions included the introduction of full time operators (3.7%); allowing an operator to be present from start of the process with a witness (3.7%) and finally,

³⁶ ABE states that videotaped interviews have become the preferred method of hearing children's evidence in criminal (and civil) proceedings and operators are now encouraged to video interviews with child witnesses whenever possible.

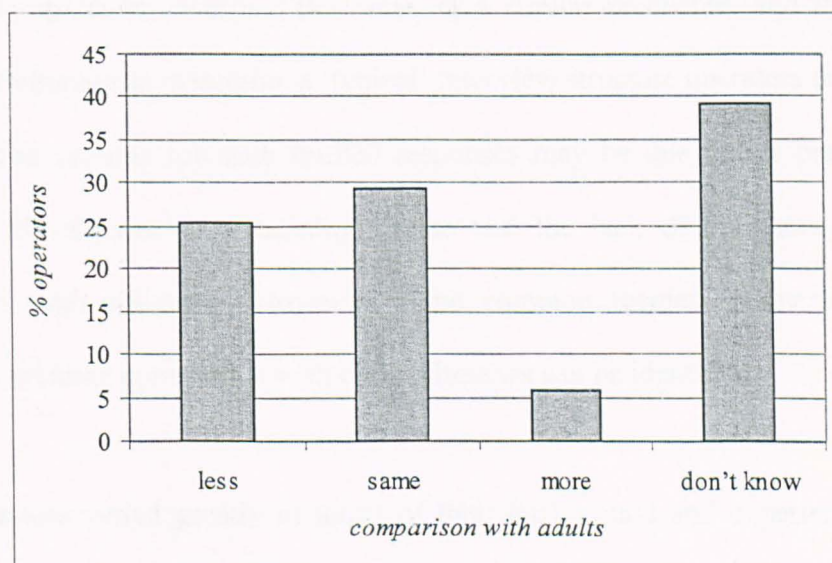
providing more careful interviewing prior to decision for composite (3.7%).

*Which features do operators think that children appear to be particularly **good** or **poor** at remembering?*

(*n*=40) The top three features children operators reported children were good at remembering were: Hair (62.5%); Clothing (32.5%); moles, lines, scars (15%). In terms of facial features, apart from Hair, only a few operators mentioned other facial features children were good at remembering, these were: Eyes, Nose and Face (10%, 7.5% and 5% respectively). (*n*=39) The top three features operators reported children were poor at remembering were: Nose (43.6%); Mouth (35.9%); and Ears (33.3%).

Do operators think that composites produced by children are less, the same or more accurate than those produced by adults?

(*n*=51) Figure 5.7 shows operators' responses.



*Figure 5.7: Operators responses to the question "Are composites constructed by children more, the same, or less accurate than composites constructed by adults?" in Study 1 (*n*=51)*

From Figure 5.7 it can be seen that the majority of operators did not know how the accuracy of children's composites compared to adults. There was a group of operators who believed that children's composites were *more* accurate than adults (5.9%). These

operators were constructing between 11.1% and 50% of their composites with child witnesses a year and 1 operator had constructed a composite with a 6-year-old.

5.4 DISCUSSION OF STUDY 1

5.4.1 DISCUSSION OF STUDY 1 RESULTS

An analysis of data from the questionnaire survey has highlighted a number of issues concerning the interviewing of children in order to create facial composites. This section will discuss these issues in relation to the aims of Study 1.

Operators' current practice in creating composites with child witnesses (aim (i))

At the end of the questionnaire survey operators were asked to describe their preferred method of producing a facial image with a child. Unfortunately, responses were fairly limited. No operators described the same, or a similar procedure, and there was not enough information to determine a 'typical' interview structure operators are using with children. The reasons for such limited responses may be due to the practical issues including the absence of a training course and the lack of operators constructing composites with children. However, some common themes in operators' current practice in creating composites with child witnesses can be identified.

First, operators varied greatly in terms of their background and experience: from the length of time they had been an operator to the numbers of composites they were constructing with witnesses (both adults and children). There were operators who stated that there had been no need for them to construct composites with children although current crime figures show that crimes are being committed against children (e.g. Creighton, 2001; Harrington and Mayhew, 2001; Kilsby 1999). Reasons for not conducting E-FIT interviews with children included statements about child witnesses

which were not supported by any appropriate evidence. Operators' having little practice interviewing children only appears to contribute to the ambiguity surrounding children's abilities to construct composites.

There were also large variations in the age of the children operators have interviewed in order to construct a composite. For most operators the youngest child they had interviewed was 12-years of age and no operator reported having constructed a composite with a child less than 6-years of age.

The majority of operators had received official E-FIT training in the form of a two week course, although most attended courses approximately 5 years ago and some, over 10 years ago. There were operators who had not attended training courses and were self taught or taught through other users. In addition, only a small percentage of operators reported updating their training at regular intervals. Very few operators attended other courses on conducting interviews and again, those operators who had, attended these courses years ago.

Operators reported spending varying amounts of time with child witnesses. ABE states that "It is not possible or desirable to put forward an ideal duration for an interview, though many interviews in practice last around one hour. However, rather shorter times may be necessary for developmentally younger children with limited attention spans, while the oldest children may be comfortable with an interview which lasts longer" (Home Office, 2002, p.35). However, most operators reported spending the same amount of time with a child witness as with an adult witness and on average, operators reported spending just over two hours with a child witness.

In the composite section of the interview, operators were using E-FIT's feature searching component with most frequency with child witnesses, followed by changes in the paint package. The majority of operators reported using each construction component with the same frequency with child and adult witnesses. A number of operators also specifically stated they use the same method with a child and an adult witness.

Operators reported using additional materials with adult and child witnesses, such as shape and size charts, lists of descriptors and a pen and paper. Operators also reported showing E-FIT's description boxes to adult and child witnesses in an attempt to prompt the witness for a more detailed description and to focus and involve witnesses. Previously operators were instructed during training not to show description boxes to witnesses. Regardless of this advice many operators did show the description boxes to witnesses anyway and training has now changed to allow this (C. Clark, personal communication, 12th October, 2004).

Finally, a number of operators stated that they made sure they completed the description interview with a child witness before turning the computer on. A few operators stated that the reason for doing this was that children focused too much on the computer when it was turned on (and presumably not enough on their description).

Operators' experiences with, and opinions of, child witnesses (aim (ii))

There appeared to be no consistent relationship between an operators' experiences and the age of child they believed it was possible to interview. For some operators there seemed to be an overall opinion that individual children vary regardless of age, which is consistent with research that age is only one factor of children's performance and other factors can be equally, if not more significant (Davies, 1996). However, the majority of

operators appeared to focus on a single age at which children may be interviewed and some operators reported refusing to interview children simply on the basis of their age, for example, one operator reported that “I once refused to construct an image with a 6-year-old girl as I considered her too young”. Most operators would be willing to interview a child younger than they have had experience of interviewing. For the majority of operators this age was 10-years although some operators stated that they would be willing to interview a 5-year-old.

On the basis of their experience with and opinions of child witnesses the majority of operators changed their interview procedure when interviewing children. One of the main changes included simplifying and modifying their language (e.g. providing more explanations; changing questions and using less formal wording). Operators described that in the free recall section of an interview adults had superior language abilities when compared to children and as such, they provided more detailed free descriptions. This is consistent with the research described in Section 2.2.2: that children’s free descriptions tend to contain little detail (e.g. Lampien and Smith, 1995; Schwartz-Kenney, Bottoms and Goodman, 1996) and that the quantity of free descriptions increases with age (for a review see Davies and Westcott, 1999). Adults’ free descriptions can be up to three times as long as 5- to 6-year-olds (Leippe *et al.*, 1991). As such, operators also reported using open questions with adults, whereas they described that children responded well to specific questions, which is again consistent with research described in Section 2.2.2 (e.g. Dent, 1992; Orbach and Lamb, 2000). As described, operators are currently using materials with witnesses in order to prompt for a more detailed description and in the final part of the questionnaire a number of operators stated that they would like to receive a set of guidelines or prompts for use with witnesses.

Operators reported changing the time spent interviewing a child witness, for example, increasing time spent and being more patient; shortening the interview; importance of breaks. The majority of operators are of the opinion that the composite interview process is not too long for adults but is too long for children. Operators who thought that the process was too long for children stated that the whole process should be an hour, at maximum. This is consistent with the guidance provided in ABE that interviews in practice last around one hour.

Operators also mentioned that they would like to see the introduction of full time operators, and to have the opportunity to be present from the very start of the process with a witness. A number of operators stated they would like the opportunity to interview a child without an appropriate adult present. Walker (1999) stated that the presence of an appropriate adult may not be as helpful as might be expected. For example, a parent may be tempted to answer on the child's behalf and may find parts of the interview distressing. The child may rely on a parent for information and may try to alleviate the parents' distress by saying as little as possible about events which may be central to the interview³⁷.

On the basis of their experience, operators were also asked to state the facial features they believed children were good and poor at remembering. The top facial features operators reported children were good at remembering (Hair, Eyes, Nose and Face) and poor at remembering (Nose, Mouth, Ears and Eyes) appeared to be generally consistent with the feature saliency research conducted with children by Prior (1996) and Davies *et al.* (1989) described in Section 3.4.1.

³⁷ An appropriate adult does not have to be present for witnesses any more, only for suspects (C. Clark, personal communication, 11th August 2004).

Finally, as many operators have had little experience with child witnesses there were a lot of “don’t know” or “unsure” responses throughout the questionnaire. For example, the majority of operators stated “don’t know” when asked whether they thought that the process of constructing a composite changes a child’s memory of a suspects face, or whether this was more problematic for children than for adults. Another example of this is that operators did not know how the accuracy of children’s composites compared to adults. However, there was a group of operators who believed that children’s composites were actually more accurate than adults, and importantly, these operators were constructing a number of composites with child witnesses.

Practical issues faced by operators when obtaining verbal descriptions and computerised facial composites from child witnesses (aim (iii))

The first issue operators appear to face is the lack of a standardised training course, although ABE states that composite operators may “require additional training and support” (Home Office, 2002, p.10). The majority of operators had attended training courses on interviewing adult victims and witnesses in order to construct a composite image. Bruck, Ceci and Hembrooke (2001) stated that it was not sufficient that professionals be provided with training: the training programs themselves must contain the most up to date and relevant materials. However, operators who had attended such courses did so a number of years ago and very few operators regularly updated training. Additionally, the majority of operators stated that the courses they attended were approximately two weeks long. Research has shown that good interviewing skills cannot be taught in such a short period of time and that interviewers often fall back into their interviewing techniques prior to the training course after a period of time (Memon, Bull and Smith, 1995). Other research on interviews for suspect child abuse victims has shown that there are difficulties in making interviewer training successful (e.g. Westcott, Kynan and Few, in press) and that there is a need for police officers to make

time for rehearsal of skills learnt at training with feedback of the trained skills (e.g. Westcott and Kynan, in press). Clark (2002) has also presented evidence that not all operators rigidly adhere to the E-FIT training they have received in their real world police work.

The questionnaire highlighted the absence of a training course specifically to provide guidance on constructing composites with children, and a number of operators stated that they would like to receive some training for constructing composites with children. Operators did describe seeking informal training and material, consistent with research which found that in some countries a number of lawyers, like police officers and social service personnel before them, are now taking the time to learn about relevant developmental issues so that they can question children appropriately (Bull, 2001). However, a number of operators reported that they were self taught and very few operators reported attending other courses offering training for interviewing children. On the one hand, there were some operators who appeared to seek out training but on the other hand there were some operators who had not attended any courses.

A second issue which was raised throughout all sections of the questionnaire is that of time: the amount of time spent interviewing child witnesses and related to this, children's concentration levels. Operators reported spending varying amounts of time with child witnesses, and were often unsure whether the process was too long for children. Some operators increased the amount of time they spent with a child, other operators shortened the interview and many operators mentioned the importance of breaks. Research and practice has shown that interviewers need to: allow extra time for child witnesses to respond; slow down their speech rate; and include more breaks and pauses. ABE does not provide a final statement about the duration of an interview as it

varies between individuals, as described earlier in this section.

Although the majority of operators reported spending the same amount or a shorter amount of time with a child as with an adult witness, in other parts of the questionnaire operators appeared to be concerned about children's concentration levels and attention spans. The issue of children's concentration was first mentioned when operators were asked what the decision of whether to interview a child was based upon. Although not provided as an option, a few operators reported that the concentration/attention span of a child was a determinant of this decision. Following on from this, concern with children's concentration levels (when compared to adults) was the most frequently reported problem in both the construction and description stages of the interview (e.g. children being easily distracted). Changes surrounding the concentration levels of children were also the second most frequently mentioned change when operators were asked in what ways they treated children differently to adults on the basis of their experience.

Language is a third practical issue that emerged from operators' responses. As with concentration, language was mentioned throughout all sections of the questionnaire. Language was the second most reported problem by most operators in both the construction and description stages of the interview. It was also the most frequently mentioned change when operators were asked in what ways they treated children differently to adults on the basis of their experience. Operators reported that in the composite stage of the interview the longest time was spent on feature searching, perhaps suggesting that better initial descriptions of features may assist in this stage. As described throughout Chapters 2 to 4, difficulties with language may pose problems at both the level of the witness and the level of the operator. Operators detailed changes to

avoid problems at both of these levels, including simplifying and modifying their own language, providing more explanations, changing questions and using less formal wording with children. One operator also mentioned observing child witnesses for their non-verbal communication.

Operators were asked specifically how the process compares to adults. Although they described adults as having better language abilities and better concentration levels than child witnesses, most operators responded that they treated children in the same way as adults, although the research described in Chapter 2 indicates that children require prompts (e.g. Orbach and Lamb, 2000) and/or special interview techniques such as the NE procedure (Saywitz and Snyder, 1993) or non-verbal props/prompts (e.g. Pipe *et al.*, 2002) in order to obtain the greatest quantity of accurate information.

Perhaps in an attempt to reduce the problems of language and concentration, operators reported using additional materials with adult and child witnesses. Various forms of materials were described although similar materials were reported to be used with adults and children. Operators did not appear to regard E-FIT's description boxes as a material: only some operators reported using them when asked about materials. However, when specifically asked, approximately half of operators reported using these boxes to witnesses to prompt for a more detailed description (relating to the issue of language) and to focus and involve witness (relating to the issue of time/concentration). Apart from E-FIT's description boxes, there was a lack of standardised format within and between forces and to date no research has been conducted into materials used with witnesses. Finally, a number of operators appeared to be eager to receive a set of guidelines or prompts for use with witnesses.

The majority of operators stated the current procedure used to obtain descriptions and composites from children was not optimal and almost all operators reported problems when interviewing children (in both the description and composite stages). The most frequent changes operators would like to see related to the issues already identified: first, training or a set of guidelines or prompts in order to tailor the composite process to suit the needs of a child witness (relating to the issue of language), and secondly, some way to shorten the whole process (relating to the issue of time).

5.4.2 LIMITATIONS OF STUDY 1

The main limitations of Study 1 are those associated with postal surveys. The proportion of non-returnees was high (as is usual with postal surveys). Although 277 operators were sent questionnaires, this number is an overestimation of the number of composite operators in the UK. Additionally, the author could not follow up any operators who had not completed questionnaires as the names and addresses of respondents remained confidential to Aspley Ltd. There is no certainty that the non-returnees are not constructing E-FITs or not constructing E-FITs with children.

The author also had no control over respondents' interpretation of the questions. Although the pilot removed any ambiguous questions, and the author's personal details were given to the operators for any comments or queries, there was still a set of responses to one question which remained ambiguous and could not be used as respondents answered in an inappropriate way.

Although face to face interviews were conducted with a few operators, a postal survey was the only viable option for obtaining responses from numerous operators across the UK. The use of a postal survey also provided respondents with the option to remain anonymous and therefore perhaps the survey obtained more honest answers from

operators.

5.4.3 SUMMARY: IMPLICATIONS OF STUDY 1 AND FUTURE RESEARCH

A number of implications for future research arise. Some issues can be isolated and considered in terms of the current thesis, and other issues can be highlighted as future directions for research which is beyond the scope of the current thesis. The latter will be considered first.

Providing a standardised training course, specifically on producing composites with children, is an important area of future policy and practice development. Future research questions may address what type of training would be most beneficial to operators and how to sustain skills and knowledge obtained in training. Further, such research may examine any effect of training on the other issues of language and time raised in the questionnaire responses.

In addition to the questionnaire responses, speaking to operators at conferences and informal interviews with operators frequently raises the issue of operators feeling they are an isolated group. Operators were not mentioned at all in MoGP (Home Office, 1992). An important first step has been made as operators are mentioned in the government's new guidance (ABE, Home Office, 2002). However, this guidance only includes one paragraph relating to operators. As a result, many operators do not consider themselves as 'interviewers' and therefore do not follow the government's guidance. Perhaps the introduction of full time operators (which was suggested by operators in the questionnaire) would lead to a more established and experienced group which would in turn address some of the main issues with child witnesses. Indeed, another questionnaire of operators following the publication of ABE may be useful to determine exactly how many operators consider themselves as 'interviewers' rather than

just operators.

Future research beyond the scope of this thesis could evaluate the existing materials operators are using with witnesses in terms of their effect on the quantity and accuracy of descriptions obtained.

Questions which may be considered in terms of the current thesis include those relating to the interview process between an operator and a child witness. First, **language**: questions surrounding the language problems of both operators and child witnesses. Secondly, **time** and **concentration**: these raise questions about whether the whole interview process could be shortened, how distractions could be reduced in order to maintain children's concentration and attention levels. The use of materials by operators in order to address these issues may also be examined in more detail. Although the existing materials operators use are not standardised, approximately half of operators report using E-FIT's description boxes both to prompt and focus witnesses, and these boxes are standardised.

In terms of the current thesis, it is also important to note that no 'typical' interview structure for use with child witnesses may be used as a comparison, or control, interview structure in future experimental studies.

In general the survey has identified the need for further research that will determine effective interviewing techniques for producing composites with children, and it is this need which has directed the current programme of research.

CHAPTER 6: STUDY 2

CHILDREN'S VERBAL DESCRIPTIONS OF UNFAMILIAR FACIAL COMPOSITES

This chapter describes Study 2, an exploration of the content and quantity of children's verbal descriptions of unfamiliar faces. Study 2 comprised two experiments: Experiment 1 describes adult participants' selection of a set of facial composite images to create a set for use with child participants in Experiment 2. Experiment 2 involved the collection and investigation of children's descriptions of the unfamiliar facial composite images.

6.1 INTRODUCTION

Chapters 2 to 4 emphasised the importance of a witness being able to provide a clear and detailed facial description of a suspect when producing a composite. Studies which have examined children's facial descriptions in terms of feature saliency have shown that not all facial features are remembered equally and that Hair has the highest feature saliency followed by Eyes for children (Davies *et al.*, 1989; Prior, 1996). It is these salient features which are the basis for subsequent recognition (Shepherd *et al.*, 1981). For example, Want *et al.* (2003) found that children's facial recognition skills were faster and more accurate for outer facial features than for inner features.

As described in Chapter 3, the majority of facial recall research has concentrated on collecting adults' terms for describing faces. Indeed the feature index for the E-FIT composite system was built using terms collected from a population of adult participants (Shepherd and Deregowski, 1981).

Study 1 identified practical problems faced by composite operators when obtaining facial descriptions from children. As described in Chapter 2, language development places limits on the extent to which children can communicate their experiences and may limit their recall skills even if the information is available. Therefore it is usual to find that young children's reports are less likely to be complete and/or accurate than older children's or adults' reports (e.g. Bull, 2001; Cunningham and Oden, 1986; Levine and Shevall, 1985; Pedelty *et al.* 1985). In particular, children (especially younger children) often have difficulty responding to free recall questions. There is a consensus among researchers that although children's free descriptions can be accurate they tend to contain little detail (Goodman and Reed, 1986; Lampien and Smith 1995; Leippe *et al.*, 1991) and that the completeness of free descriptions tends to increase with age (see review by Davies and Westcott, 1999). It is therefore necessary to ask witnesses, and in particular child witnesses, more questions in order to trigger memories and elicit the fullest possible accounts (Orbach and Lamb, 2000). Research consistently shows that prompting witnesses increases the amount of information provided (Davies and Westcott, 1999).

The data from Study 1 indicated that a large proportion of operators (just over 40%) were using E-FIT's adult feature index to prompt children to provide more detailed descriptions. However, children have difficulty understanding the meaning of words used by adults, they may also have their own language and may use terms differently to an adult (Saywitz, 1989; Saywitz *et al.*, 1990). Children may not correct an interviewer's erroneous interpretation which could prove to be particularly problematic as operators are required to interpret and 'translate' witnesses descriptions (adults' and children's) to input them into the E-FIT's (adult) facial feature index (E-FIT 3.1 for Windows: On-line help file).

Missing in the literature described in Chapters 2 to 4 is a detailed examination of the content of children's descriptions of unfamiliar faces at different ages. Previous work has been conducted to examine children's vocabulary for sexual experiences (Volbert and Van der Zanden, 1996) and their use and understanding of legal terms (Saywitz *et al.*, 1990). However, this has not been extended to the area of facial recall and descriptions.

Gathering knowledge about the language and terms children use to describe unfamiliar faces is important for informing research on how effectively children can produce verbal descriptions and computerised facial composites of an unfamiliar face. Documentation of the specific language and terms used by children could assist in making the communication process between a witness and a police operator more effective, by aiding the development of age appropriate vocabulary as a possible prompting technique for producing composites with children and assisting operators to interpret children's descriptions accurately.

In order to explore the content of children's descriptions of unfamiliar faces in detail, an appropriate methodology is required and the application of research which has examined children's interpretations of facial expressions may be of use. The free labelling method has been used since the early twentieth century with children (e.g. Gates, 1923) and typically involves an experimenter showing a child a still photograph of a facial expression and asking the child "how is this person feeling?". This method avoids the criticisms associated with the more commonly used method of forced choice labelling of still photographs of facial expressions, as it does not force children to choose a label they would not have thought of spontaneously, or may have even rejected

(Russell, 1994). Free labelling allows an examination of children's spontaneous use of labels and what they reveal about children's descriptions of faces rather than simply categorising them as correct or incorrect. Such a technique may be adapted in the present study to establish the language and terms children use to describe unfamiliar faces.

Finally, in order to gain the greatest and most detailed content from children's descriptions, the present study employed the use of a new interviewing technique. Existing studies of children's descriptions, which focus on completeness, accuracy and feature saliency, use verbal prompts that encourage featural facial descriptions (e.g. "what can you tell me about the eyes?", "what can you tell me about the nose?" and so on). With the aim of encouraging children to include configural information³⁸ in their descriptions, the present study employed the use of a 'comparison' description interview technique, where children were required to compare two composite faces.

6.1.1 RESEARCH AIMS OF STUDY 2

The main aim of Study 2 was to investigate the *content* of children's descriptions of unfamiliar faces (i.e. what words children use to describe faces), in particular to:

(i) establish the language and terms children use to describe unfamiliar faces and to compare the language and terms children use with an existing database of E-FIT (adult) terms for unfamiliar faces. This aim was treated as exploratory and no hypotheses were stated.

³⁸ Configural information is defined as sensitivity to second order relations (representations of the spatial relations among the individual features), for example "his eyebrows are almost touching his nose" (Maurer *et al.*, 2002; Pellicano and Rhodes, 2003).

An additional aim of Study 2 was to investigate the *quantity* of children's descriptions of unfamiliar faces in order to replicate existing findings in the literature on children's event reports, in particular to:

(ii) *investigate the effect of age on the quantity of children's descriptions of unfamiliar faces*. It was hypothesised that older children would provide statistically significantly more descriptions than younger children.

(iii) *explore the effect of description type (free versus prompted) on the quantity of children's descriptions of unfamiliar faces*. It was hypothesised that prompting would lead to a statistically significant increase in the quantity of children descriptions.

A further aim of Study 2 was to investigate the effect of a *comparison description task*, in particular to:

(iv) investigate the effect of a comparison description task on the quantity of children's descriptions of unfamiliar faces. This aim was treated as exploratory and no hypotheses were stated.

(v) investigate the effect of a comparison description task on the configural content of children's descriptions of unfamiliar faces. This aim was treated as exploratory and no hypotheses were stated, although it was expected that all children would provide some configural descriptions.

Finally, for all of the aims, the effect of different age groups was explored, for which no hypotheses were specified unless otherwise stated.

6.2 EXPERIMENT 1: DEVELOPMENT OF FACIAL STIMULI FOR EXPERIMENT 2 (ADULT PARTICIPANTS)

For both Experiments 1 and 2 a set of facial stimuli had to be created. Facial composites built using the E-FIT system were used (rather than photographs of faces) in order to provide an established and objective documentation of E-FIT's (and therefore adults') description terms for all of the features and feature elements within each composite. Furthermore, analyses of the data could directly compare the descriptions obtained from child participants to the E-FIT (adult) description terms for each composite.

In order to reduce any subjectivity in the creation of these facial stimuli, and to obtain a smaller set of stimuli that contained a wide range of facial features and could be described with some consistency, adult participants were asked to take part in a task that would reduce the number of composite images for use in Experiment 2 with child participants.

6.2.1 METHOD

Design

A mixed factorial design was used. 'Description' was a within participants factor and had two levels (Free and Prompted). 'Face Order' was a within participants factor, and had four levels (the First, Second, Third and Fourth composite face shown to participants). The dependent variable was the descriptors participants used to describe the facial images.

The 16 facial composite images were pseudo-randomly assigned to participants with the constraint that no participant viewed the same set of composites. The order of description (free followed by prompted) was fixed so that the prompts did not interfere with the free descriptions.

Participants

Eight adults (four males and four females) were drawn from students, staff and visitors to The Open University. They were recruited informally and were not paid for their participation. Their ages ranged from 21-years to 59-years (mean=37-years, 5-months). All participants spoke English as their first language.

Materials

Sixteen greyscale, two-dimensional, computer generated facial composite images of Caucasian adult males were created using E-FIT (version 3.1a for Windows). The set of images was constructed by the author using the full E-FIT feature index as a checklist in order to create images which covered as wide a range of different exemplars for each feature element as possible (e.g. all of the different exemplars for the feature element Face Shape were used). No composites shared the same example of a feature (i.e. although two composites may have both had the exemplar 'oval' for the feature element Face Shape, two different oval face shapes would be used). The images were not based on any particular persons and were therefore unfamiliar to participants. Each composite image was edited using Adobe PhotoShop 5.0 in order to produce 16 composites of a standardised size of 10cm by 15cm. The composite images were printed in greyscale at 72 d.p.i resolution (E-FITs maximum output). One composite image was printed per A4 page. The images can be seen in Appendix IIIa.

In addition to the facial stimuli, three-page description forms were provided to enable participants to complete free and prompted written descriptions of the composite images.

Procedure

1. Free written description of composite image 1

Each participant was provided with one facial composite image and one description form. Participants were told they could view the composite image for the duration of the task. On the first page of each description form the participants were asked to “provide a description of the person”. At the bottom of the first page was the written instruction to participants “when you have finished your description please turn to the next page”.

2. Prompted written description of composite image 1

On the next two pages, participants were instructed “for each of the following features provide as full as a description as you can, even if it means providing details you described earlier”. The pages contained prompts related to E-FIT’s seven main features (Hair; Face; Eyes; Eyebrows; Nose; Mouth and Ears) and seven of E-FIT’s feature elements (Hair Style; Hair Length; Hair Style; Face Shape; Chin, Cheeks, Forehead).

The procedure described above for the free and prompted written description of composite image 1 was repeated for a further three composite images.

All adults were interviewed individually. At the end of the task, participants were thanked and de-briefed.

6.2.2 RESULTS

Treatment of Raw Data

In order to reduce the set of facial images from 16 to 10, for use in Experiment 2 with child participants, the data was treated in the following way. Adults’ written descriptions provided in both free and prompted written sections of the interview were combined and any repeated terms removed. For each composite image, descriptions were provided by two different adults. The two sets of descriptions for each image were compared for matching descriptors by tabulating them against E-FIT’s feature element

categories adult participants were prompted for. Table 6.1 shows an example of this tabulation.

E-FIT Feature Element	Composite Image 2	
	Participant 1	Participant 2
Hair Length	<i>short</i>	<i>short</i>
Hair Type	<i>curly, full</i>	<i>curly, soft</i>

Table 6.1: Example coding of adults' descriptors in Experiment 1

In Table 6.1 matching descriptors are shown in italics. For example, participant 1 *and* participant 2 used the descriptor “short” to describe the feature element Hair Length for composite image 2. This counted as 1 matching descriptor for image 2. The total number of descriptors used by adults which *matched* for each composite face was calculated. Table 6.2 summarises this information.

Composite Image	Total n°. of matching descriptors
1	11
2	6
3	9
4	9
5	3
6	9
7	9
8	11
9	12
10	7
11	3
12	8
13	11
14	4
15	8
16	6

Table 6.2: Total number of matching descriptors provided by adults for each composite face in Experiment 1

Table 6.2 shows that the six composite images for which adults' responses contained the fewest matching descriptors were: 2, 5, 10, 11, 14, and 16. Therefore, these composite faces were removed in order to leave ten composites which could be

described with some consistency for use as the facial stimuli with child participants in Experiment 2.

6.3 EXPERIMENT 2: CHILDREN'S VERBAL DESCRIPTIONS OF UNFAMILIAR FACIAL COMPOSITES

Experiment 2 involved the collection and investigation of children's descriptions of the ten unfamiliar facial composite images produced in Experiment 1. On the basis of the existing developmental literature described in Chapter 2, and operators' responses to the questionnaire in Study 1, the children were equally divided into three age groups of 6-, 8-, and 10-year-olds. They were asked to complete two tasks: first, a Description Task asked children to provide free and prompted verbal descriptions of two composite images (presented individually); secondly, children were asked to provide a Comparison Description, requiring a comparison of the two composite images.

6.3.1 METHOD

Design

A mixed factorial design was used. 'Age' was a between participants factor, and had three levels (6-, 8-, and 10-years). For the first task, 'Description' was a within participants factor and had two levels (Free and Prompted), and 'Face Order' was a within participants factor, and had two levels (Primary and Secondary). All participants completed the second task of a 'Comparison Description'. Children's descriptions consisted of a number of *descriptors*. The dependent variables were the content and quantity of descriptors used.

The ten facial composite images were pseudo-randomly assigned to participants with constraints so that no participant viewed the same pair of composites.

Participants

Thirty children (16 boys and 14 girls) were drawn from one lower school and one middle school in Bedfordshire (UK) and were equally divided into three age groups. The first group comprised five boys and five girls from Year 1, aged between 5 years 10 months and 6-years 9-months (mean=6-years 4-months). The second group comprised five boys and five girls from Year 3/4, aged between 7 years 11 months and 8-years 9-months (mean=8-years 3-months). The third group comprised six boys and four girls from Year 6, aged between 10-years 4-months and 11-years 2-months (mean=10-years 10-months). Parental and participant consent was gained for each child and children were not paid for their participation. All children spoke English as their first language.

Materials

Facial Stimuli

Ten greyscale, computer generated facial composite images of Caucasian adult males created in Experiment 1 were used as facial stimuli.

Practice Stimuli

Two greyscale images of houses were used as practice images³⁹. The practice images were also edited using Adobe PhotoShop 5.0 to produce images of a standardised size of 10cm by 15 cm (see Appendix IIIb).

An audio recorder with a built in microphone (Sony TCM-4ODV) was used to record children's responses. An instruction sheet was used to ensure consistent prompting of children, and to ensure guidelines for interviewing children (Home Office, 2002) were followed as far as possible.

³⁹ Houses were used as a practice stimuli as they include both featural and configural information (Tanaka and Farrah, 1993).

Procedure

A pilot study was conducted with a 6-year-old girl to test the facial composite images, the instructions for child participants, the practice interview and the feasibility of the procedure as a whole. Pilot data are not included in the results.

All children were tested individually, in a quiet area of the school free from distractions.

The general procedure followed the instruction sheet and is summarised below:

1. Introduction

Each child was introduced to the interviewer who explained the purpose of the study. Children were informed that at the end of the interview they would return to their classroom and that, if at any time during the procedure they wanted to go back to their classroom they could tell the interviewer. During the introduction it was emphasised to the children that there were no right or wrong answers and that it was acceptable for them to say “I don’t know” to any of the questions they were asked.

2. Practice images

On the basis of the research described in Chapter 2 and guidance from Home Office (2002), a practice interview was included to further establish rapport and to inform participants of each stage of the interview and the level of information required of them.

Free description of practice image 1

Children were asked to view one image of a house, and with the image remaining present⁴⁰, children were first asked the open ended prompt “Tell me everything you can about the house”. Throughout the free description section facilitators were provided by the interviewer (e.g. “OK”, and “anything else?”)

⁴⁰ In the current study the stimuli remained present throughout the interview as the main aim was to gather the verbal terms children use, not to test their memory.

Prompted description of practice image 1

When children had described the house in as much detail as possible, they were then provided with specific featural prompts relating to the sizes and shapes on the houses (e.g. “Tell me what shapes you can see on the house”).

Free description of practice image 2

When children had finished describing the first house they were asked to view an image of a second house. The procedure described above for the free description of practice image 1 was repeated.

Prompted description of practice image 2

The procedure described above for the prompted description of practice image 1 was repeated.

Comparison description of practice images 1 and 2

When children had described the images separately in as much detail as possible, they then viewed both practice images together. Children were asked “Do you think the houses are the same or different?” and “Why do you think that?”. Children were then prompted to compare features and the location of the features on both of the practice images (e.g. “What things about the houses are different and what things are the same?”). Children were also provided with an example if needed (e.g. “Something that is the same is that they both have windows but something that is different is that the windows on this house are small and the windows on this house are big”). This was to enable children to practice providing configural descriptions.

3. Facial stimuli

After the practice task, children were shown the facial composite images. Each image remained present throughout the corresponding section of the interview and the procedure described above with the practice images was followed:

Free description of composite image 1

Children were asked to view one of the 10 composite faces. Again, the interview began with an open ended prompt “Tell me everything you can about the face”. And facilitators were provided by the interviewer throughout until children had described the face in as much detail as possible.

Prompted description of composite image 1

In order to supplement children’s free descriptions, they were then provided with specific featural prompts by the interviewer for E-FIT’s seven main features (Face; Hair; Eyebrows; Eyes; Ears; Nose; and Mouth) and six feature elements (Face Shape; Nostrils; Hair Colour; Hair Length; Hair Style; and Hair Type). If a child had not mentioned a feature they were asked “do you know what these are <point to feature>?”. If they did not provide a name they were told that the experimenter called them <state name of feature> and asked “what do you call them?” and “what do they look like?”. Children were also asked to think about the descriptions they provided for the houses to encourage them to consider shapes and sizes.

Free description of composite image 2

When children had finished describing the first composite image they were asked to view an image of a second composite image. The procedure described above for the Free description of composite image 1 was repeated.

Prompted description of composite image 2

The procedure described above for the Prompted description of composite image 1 was repeated.

Comparison description of composite images 1 and 2

When children had described both of the composite images separately in as much detail as possible, they then viewed both composite images together. Children were asked “Do you think the faces are the same or different?” and “Why do you think that?”. Children were then prompted to compare features and the location of the features on both of the composite images. (e.g. “What things about the faces are different and what things are the same?”) and again children were provided with an example if needed.

4. Closure

At the end of the interview children were thanked, their well-being checked and they were asked if they had any questions about what they had just done. Children were then returned to their classroom and were asked not to discuss the interview with other children. With the consent of the children’s parents all of the interviews were audio-taped. Finally, all results were displayed in a ‘child friendly’ format to each school at the end of the study.

Throughout the interview all children were observed closely. The number of facilitators and prompts provided by the interviewer in all sections of the interview depended on each individual child and was assessed by the interviewer. The interviews ended when a child had exhausted their descriptions or appeared to be tired or to have lost concentration. The children were allowed as long as they needed to view the practice images and facial composite images and to describe them. The majority of interviews lasted between 20 and 30 minutes and no interview exceeded 35 minutes.

6.3.2 RESULTS

This section begins with a description of the treatment of the raw data. It then considers the aims of the study. The content of children's descriptions is considered first (consisting of mainly qualitative, descriptive data which is considered both by-participant and by-feature). The quantity of children's descriptions is examined next (consisting of mostly quantitative data analysed by participant). Finally, the effect of the comparison description task on the content and quantity of children's descriptions is considered.

Treatment of Raw Data

The tape-recorded descriptions were transcribed and coded by the author. In order to allow a comparison of children's descriptions to the existing E-FIT (adult) descriptions, responses were primarily coded by tabulating responses using the E-FIT feature index. Seven main feature categories and 19 feature element categories which were used when creating the facial composite images were used⁴¹.

Descriptions provided by a child in all parts of the interview were tabulated against the E-FIT (adult) terms for each composite face. Table 6.3 demonstrates this tabulation.

⁴¹ A list of the feature element categories is provided in Table 6.6.

Feature: Eyebrows			Descriptions		
<i>E-FIT Image</i>	<i>E-FIT terms</i>	<i>Child</i>	<i>Free</i>	<i>Prompted</i>	<i>Comparison</i>
4	Arched; Thin; Lowered ends; light	1	aren't that big	(no description)	smaller; rectangles; higher; going in
		2	really straight	(no description)	straight
7	Peaked; Thick; Angled ends; Dark	1	(no description)	quite big	(no description)
		2	(no description)	like rectangly; cross; because going diagonally	just touching where nose connects onto face

Table 6.3: Example coding of children's descriptions tabulated against E-FIT description in Experiment 2

Children's responses were divided into descriptors. The total number of descriptors provided by each child in each section of the interview was calculated. For example, the description "he's got short, curly hair" would count as 2 descriptors for the main feature category Hair, or as 1 descriptor for the feature element category Hair Length and 1 descriptor for the feature element category Hair Type. "His Mouth is the same as the other one" (referring to the previous composite face viewed) would only count as 1 descriptor if a descriptor was provided for the relevant feature on the previous face viewed. "His mouth is the same shape as his lips" would only count as 1 descriptor if a descriptor had been given to the feature referred to.

Negative descriptions were scored in the following way, for example "he hasn't got a smiley mouth" would count as 1 descriptor for the main feature Mouth. All of the children's descriptions were considered 'correct' as the aim of the study was to gather children's terms and language used to describe unfamiliar faces.

In order to further categorise children's descriptions, in addition to the E-FIT feature index the following categories were included:

A *configural* category was created to code configural descriptions (as defined previously). For example “his eyebrows are almost touching his nose”.

A *non-verbal* category was created to provide information about children drawing features, children tracing features (in the air, on the table or on paper) and children pointing (to their own face, to the interviewer’s face or to the stimuli).

A *location* category was created to provide information about descriptions provided by children about a participant part of a feature. For example “his hair is curly on the *top*”⁴² would count as 2 descriptors for the main feature Hair.

Qualitative Analyses – The Content of Children’s Descriptions of Unfamiliar Faces

The Language and Terms Children use to Describe Unfamiliar Faces (aim (i))

Initially, a comparison of children’s responses with the existing measure of adults’ terms was observed⁴³. Children’s descriptors for each of the 19 feature element categories were considered in turn. This initial trawl through the data led to a number of observations. For example, although some of the children’s descriptors were consistent with E-FIT terms large quantities were not. A number of categories were identified which the children’s descriptors varied upon. These categories included: the number of descriptors children provided; the percentage of these descriptors which were E-FIT terms; the percentage of E-FIT terms children used; the number of children providing descriptions; and, the percentage of these children who used E-FIT terms. Finally, children’s descriptors which did not match E-FIT terms were considered in terms of content and quantity.

⁴² A *features mentioned* category was created to code when children mentioned a feature (without providing a description of it) (e.g. “he has a nose”) and an *other* category to include details provided about the composite faces by children which did not fall into any of the above categories. Due to the focus of this study on the terms children use to describe faces, only descriptions which could be used by E-FIT operators are considered. Therefore, these data are not considered here.

⁴³ Configural, location or non-verbal descriptions are not included here as they will be considered in a separate section.

For example, for the feature element Hair Colour the E-FIT feature index includes five terms (Light; Medium; Dark; Greying; and Grey). For the 6-year-olds, 20 descriptors were provided in total, only 2 of which matched the E-FIT terms. Only one of the five E-FIT terms were used (Grey). Nine of the 10 children provided descriptors of Hair Colour and two children used the E-FIT term. The large number of remaining descriptors could be easily mapped onto the corresponding E-FIT terms. For example, using the descriptors “blonde” for light hair, “brown” for medium or dark hair and “black” for dark hair.

From these categories a summary table was created. An example of one section of this summary table for 6-year-olds’ descriptions for the feature element of Hair Colour is shown in Table 6.4.

<i>Feature Element</i>	<i>N^o. of descriptors provided matching E-FIT terms / N^o of descriptors provided</i>	<i>N^o of E-FIT terms used by children / total possible n^o of E-FIT terms</i>	<i>N^o of children using E-FIT terms / N^o. children providing descriptors</i>
Hair Colour	2/20	1/5	2/9
	10%	20%	22%

Table 6.4: Sample of summary table for categorisation of children’s descriptions in Experiment 2

The full summary table of the data is provided in Appendix IIIc. From the summary table and from the full details of the children’s descriptions each category was coded as either High, Medium or Low. Full definitions of this coding are provided in Appendix IIIc and are summarised in Table 6.5. Working across the example in Table 6.4 from left to right: a *High* number of descriptors were provided (n=20), of which only a *Low* number matched the E-FIT terms (n=2). A *Low* number of E-FIT terms were used (n=1) out of the total possible (n=5). A *High* number of children provided descriptors of Hair Colour (n=9) but only a *Low* number of these children used E-FIT terms (n=2).

In order to establish the language and terms children use to describe unfamiliar faces it was necessary to look at the descriptors children provided which did not match E-FIT terms. The above analyses from Appendix IIIc were observed along with the remaining descriptors provided by children (both the amount of and the variation of).

From this data, a number of patterns emerged which occurred across all conditions of the interview and across all ages:

For patterns 1 to 3, children frequently described a feature element and for:

- *Pattern 1*, a large variety of descriptors were used and these descriptors could be matched to E-FIT terms;
- *Pattern 2*, children used a small variety of descriptors but as with pattern 1, these could usually be matched to E-FIT terms;
- *Pattern 3*, children used their own descriptors which could not usually be matched to E-FIT terms

In contrast to these 3 patterns, there was a 4th Pattern:

- *Pattern 4*, children not usually providing any descriptors for a feature element at all.

These patterns are summarised in Table 6.5 (categorisations of High and Low are summarised within the table).

	<i>Pattern</i>			
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
Number of children's descriptors (<i>High = over 20; Low = under 10</i>)	High	High	High	Low
% children's descriptors matching E-FIT terms (<i>High = over 60; Low = under 40</i>)	High	Low	Low	Low
% E-FIT terms used (<i>High = over 60; Low = under 40</i>)	High	Low	Low	Low
Number of children providing descriptors (<i>High = over 6; Low = under 5</i>)	High	High	High	Low
% children providing descriptors who use E-FIT terms (<i>High = over 60; Low = under 40</i>)	High	Low	Low	Low
Number of remaining descriptors (<i>High = over 20; Low = under 10</i>)	N/A	High	High	Low
Variation of remaining descriptors (<i>High = over 20; Low = under 10</i>)	N/A	High	Low	N/A

Table 6.5: Pattern categorisation of children's descriptions in Experiment 2

Based on children's descriptors each feature element was categorised in terms of the Patterns 1 to 4 as detailed in Table 6.6.

<i>Feature element</i>	<i>Pattern</i>	<i>Notes⁴⁴</i>
Hair Length	1	
Hair Style	3	
Hair Colour	2	
Hair Type	3	6-year-olds - provided a lower number of descriptors; a lower number of children providing descriptors
Face Shape	2	
Eyebrow Shape	3	10-year-olds - a lower number of children providing descriptors.
Eyebrow Size	4	
Eyebrow Colour	4	
Eyes Shape	3	
Eyes Colour	2	6- and 8-year-olds - provided a lower number of descriptors; a lower number of children providing descriptors
Eyes Size	2	6-year-olds - provided a lower number of descriptors. 6- and 8-year-olds - a lower number of children providing descriptors
Nose Tip	4	
Nose Size (length and width)	2	
Mouth Shape	2	6-year-olds - provided a lower number of descriptors; a lower number of children providing descriptors

⁴⁴ Includes general notes about the *numbers* of descriptors and children, and different patterns for individual age groups.

<i>Feature element</i>	<i>Pattern</i>	<i>Notes⁴⁴</i>
Mouth Size (Width)	4	
Mouth Lips	2	
Ears Size	2	6-year-olds - a lower number of children providing descriptors. 6- and 8-year-olds - provided a lower number of descriptors
Ears Shape	4	6- and 8-year-olds - a lower number of children providing descriptors.
Ears Setting	4	6- and 8-year-olds - a lower number of children providing descriptors.

Table 6.6: Pattern categorisation for each feature element with notes for Experiment 2

This summary of data indicated that for the majority of facial features: 6-, 8- and 10-year-olds descriptors did not match the E-FIT terms; the number of E-FIT terms used by children was a low percentage of the total number of terms they could have used; and a low percentage of children who provided descriptors used E-FIT terms.

An example of a feature element categorised as each pattern is provided below:

From Table 6.5 it can be seen that Hair Length was the only feature element which was categorised as Pattern 1. The E-FIT feature index includes 5 exemplar terms for Hair Length, of which 4 were used in the creation of the facial stimuli (very short, short, long and very long). Twenty-six out of 30 children used at least 1 E-FIT term to describe Hair Length and in total, all 4 of these terms were used. When children did not use the E-FIT terms they simply used descriptors which could easily be mapped onto those terms, for example describing short hair as “a little bit of hair”.

Four feature elements were categorised as Pattern 2, including Face Shape. The E-FIT feature index includes 5 exemplar terms for Face Shape (oval, round, triangular, square, and angular). Twenty-two children used at least 1 E-FIT term to describe Face Shape and in total, only 3 different terms were used. Although only two of the E-FIT terms

were used frequently (round and oval) one child used another (square). The majority of children's descriptors focused on the terms round and oval to describe face shape, for example "not round" or "not oval" to describe angular, square or triangular faces, or "a stretched circle" or a "squeezed circle" to describe an oval face. There were a few children who provided new descriptors, for example the term "diamond" was used to describe an angular face. Children appeared to recognise that some faces were not round or oval but they did not appear to have any other descriptors for describing such face shape.

The majority of feature elements (8) were categorised as pattern 3 including Hair Style. The E-FIT feature index includes 9 exemplar terms for Hair Style of which 5 were used in the creation of the facial stimuli (Fringe, Layered, Punk, Spiky, and Crew Cut). Descriptions were also included which described hair thickness, parting and tidiness. Although a high number of descriptors were provided by a high number of children, only 7 children used E-FIT terms to describe Hair Style and only 2 different terms were used. The remaining descriptors provided by children were not as clear as the alternative descriptors provided for Face Shape, which could be easily matched to the feature index. Instead, children used their own descriptors. For example, "like an N" and "apart" to describe a centre parting; "mad" and "messed up" to describe untidy hair; "like a helmet" to describe a crew cut"; "like a loo brush" to describe a spiky hair.

Finally, 5 feature elements were categorized as pattern 4 (no descriptors provided), including: Eyebrow Colour, Eyebrow Size; Mouth Width and Ear Setting.

Quantitative Analyses - The Quantity of Children's Descriptions of Unfamiliar Faces

The Effect of Age on the Quantity of Children's Descriptions of Unfamiliar Faces (aim (ii))

The mean number of descriptors provided was calculated for each child by averaging the total number of descriptors (free and prompted combined) by the two composite faces shown. These values were then used to calculate the mean number of descriptors for 6-, 8- and 10-year olds⁴⁵. The mean number was calculated both including repetitions, from within or between description stages of the interview, and excluding repetitions. This data is shown in Table 6.7.

Age	Including Repetitions			Excluding Repetitions			Mean n°. of repetitions
	Min	Mean (S.D)	Max	Min	Mean (S.D)	Max	
6 years	12	18.05 (4.89)	25.5	10.5	17.1 (4.93)	25	0.95
8 years	12	21.2 (6.21)	31	11.5	19.11 (5.41)	28	2.1
10 years	15.5	25.4 (8.10)	37	15	22.35 (6.49)	33	3.05

Table 6.7: Mean number of descriptors provided by children in Experiment 2

Table 6.7 shows that the mean number of descriptors including repetitions increased with age. The mean number of repetitions of descriptors also increased with age. Therefore, the initial difference between age groups was reduced by removing any repetitions and only counting new descriptors provided by children. Further analyses do not include repeated data.

The data from Table 6.7 are illustrated in Figure 6.1 with standard error bars.

⁴⁵ Preliminarily analyses identified an outlier in the data (participant 18). This participant's data was been removed from the quantitative section of the results.

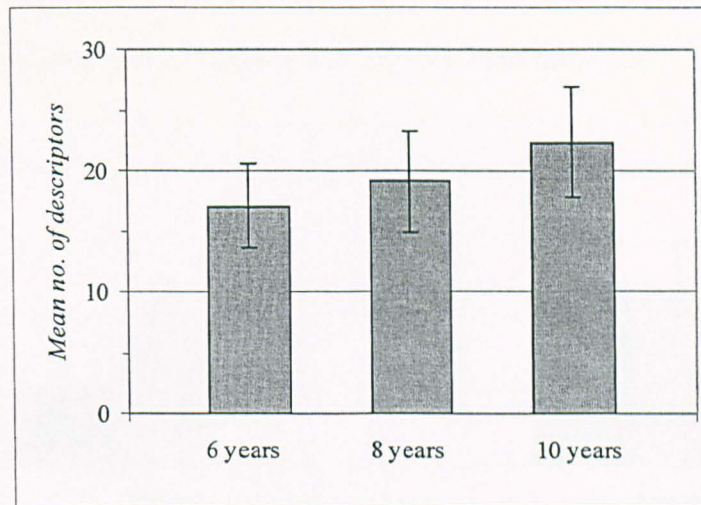


Figure 6.1: Mean number of descriptors provided by children in Experiment 2

It was hypothesised that that older children would provide statistically significantly more descriptors than younger children. A 3x2x2 mixed ANOVA design was employed. The between participants factor Age had 3 levels (6-, 8- and 10-years). The within participants factor Description had two levels (Free and Prompted) and the within participants factor Face Order had two levels (Primary and Secondary). The dependent variable was the *total* number of descriptors provided (in all sections of the interview and including all of the above categories). Results showed that although the mean scores increased with age, the main effect of Age was statistically non-significant [$F_{(2,26)}=2.190$; $p=0.066$]. Therefore the hypothesis stated above was not supported. This was due to large individual differences within each age group (as illustrated by the overlapping standard error bars in Figure 6.1).

The Effect of Description Type on the Quantity of Children's Descriptions of Unfamiliar Faces (aim (iii))

The mean number of descriptors provided in each section of the interview (free and prompted) was calculated for each child by averaging the total number of free and prompted descriptors by the two composite faces shown. These values were then used

to calculate the mean number of free and prompted descriptors provided by 6-, 8- and 10-year olds, and are shown in Figure 6.2 with standard error bars.

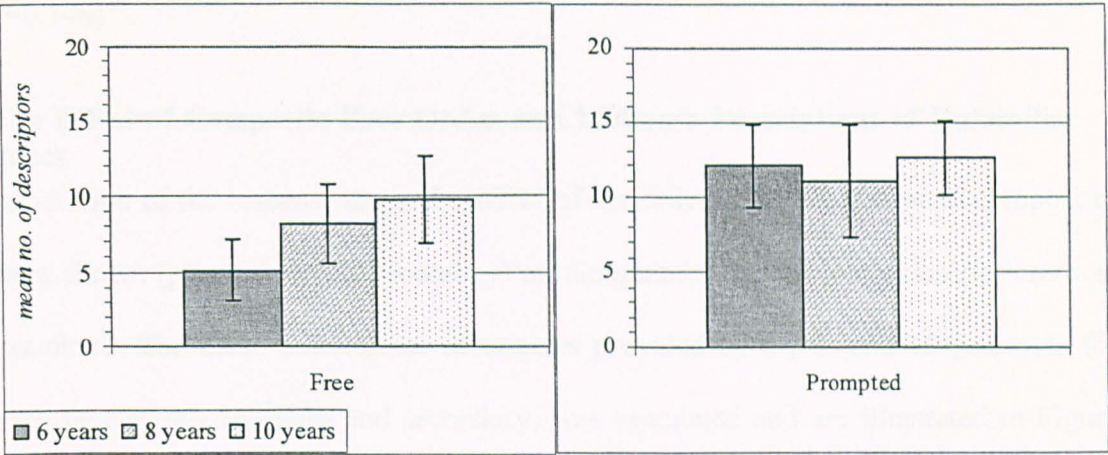


Figure 6.2: Mean number of free and prompted descriptors provided by children in Experiment 2

Figure 6.2 shows that children of all ages provided more descriptors in the prompted than in the free description section of the interview. In the free description section of the interview the mean number of descriptors increased with age, whereas in the prompted description section children of all ages provided a similar number of descriptors. The largest increase in the quantity of descriptors provided from the free to the prompted sections of the interview was for 6-year-olds, who provided a higher number of descriptors, on average, than 8-year-olds in the prompted description section. 10-year-olds provided the highest number of descriptors in both interview sections. Again, there were large individual differences within each age group (as illustrated by the overlapping standard error bars in Figure 6.2).

It was hypothesised that prompting would statistically significantly increase the quantity of children descriptions and this was supported by the results. A 3x2x2 mixed ANOVA (employed on the factors specified in the previous analyses) showed a statistically significant main effect of Description [$F(1,26)=19.842$, $p<0.0005$, partial $\eta^2=0.433$]

supporting the effect of prompting suggested in Figure 6.2. The 2-way interaction between Age and Description was statistically non-significant [$F_{(2,26)}=2.058$, $p=0.148$]⁴⁶.

The Effect of Composite Face Order on Children's Descriptions of Unfamiliar Faces

In addition to the research aims, the effect of the order in which the facial composites were shown (primary versus secondary) on the quantity of children's descriptions was examined. The mean numbers of descriptors provided by 6-, 8- and 10-year-olds for each face shown (primary and secondary) was calculated and are illustrated in Figure 6.3.

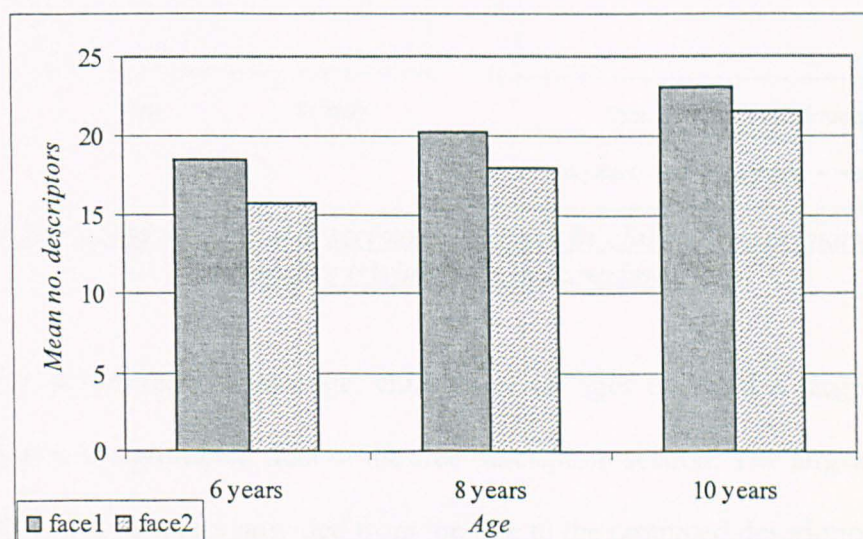


Figure 6.3: Mean numbers of descriptors provided by children for primary and secondary faces in Experiment 2

Figure 6.3 shows that for all age groups, a slightly lower number of descriptors were provided on average for the secondary composite face shown than for the primary face shown. Although, a 3x2x2 mixed ANOVA (employed on the same factors as specified previously) showed the main effect of Face Order was statistically non-significant [$F_{(1,26)}=3.514$, $p=0.072$], there was a statistically significant 2-way interaction between

⁴⁶ There was a statistically significant difference between the 6-year-olds and the 8- and 10-year-olds in the free description section of the interview [$F_{(2,27)}=3.628$, $p=0.04$].

Description and Face Order [$F_{(1,26)}=20.718$, $p<0.0005$] as illustrated in Figure 6.4. Finally, the 3-way interaction between Face Order, Description and Age was statistically non-significant [$F_{(2,26)}=1.163$, $p=0.328$].

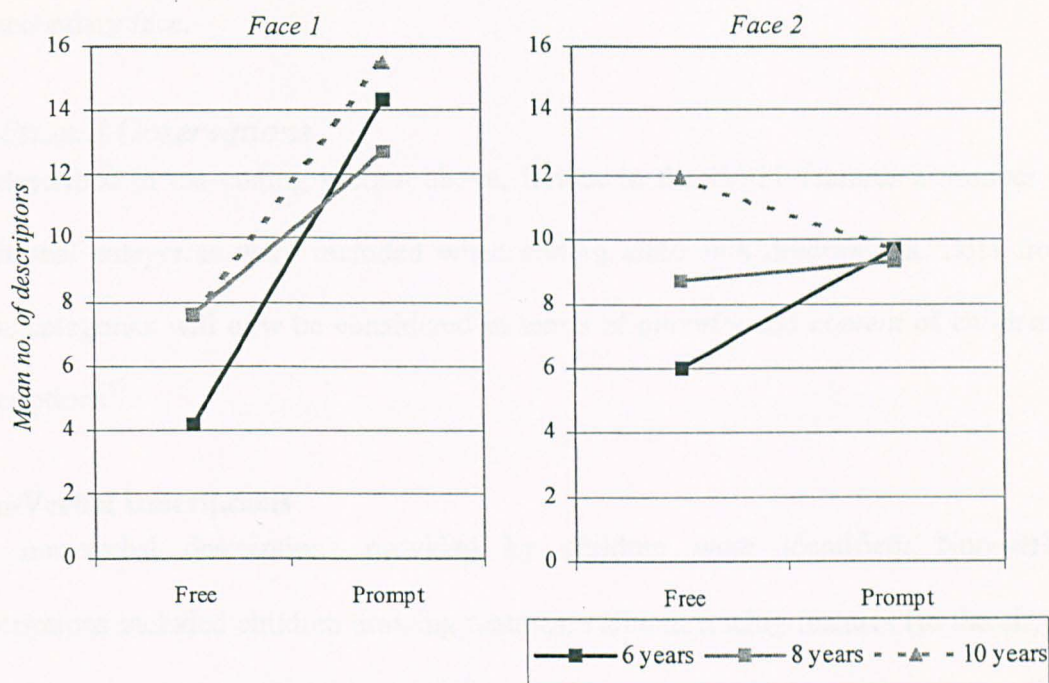


Figure 6.4: Mean number of descriptors provided by children for primary (left) and secondary (right) faces in Experiment 2

Figure 6.4 shows that, on average, children of all ages provided a larger number of descriptors in the prompted than in the free description section. The largest increase in the number of descriptors provided from the free to the prompted description sections of the interview for the primary face was for the 6-year-olds, suggesting that they may benefit most from the use of prompts. For the secondary face, 10-year-olds provided more descriptors in the free description than in the prompted description section of the interview, 8-year-olds provided a slightly higher number of prompted descriptors and again, 6-year-olds still provided a larger number of prompted than free descriptors.

When comparing the mean number of descriptors provided on the primary and secondary face, children of all ages provided a higher number of descriptors in the free

description section for the secondary composite face than for the first composite face and this increase was largest for the 10-year-olds. Children of all ages provided a similar number of descriptors in the prompted description section of the interview for the secondary face.

Additional Observations

As described in the coding section above, further to the E-FIT features a number of additional categories were included when coding children's descriptions. Data from these categories will now be considered in terms of *quantity* and *content* of children's descriptions⁴⁷.

Non-Verbal Descriptions

All non-verbal descriptions provided by children were identified. Non-verbal descriptions included children drawing features, children tracing features (in the air, on the table or on paper) and children pointing (to their own face, to the experimenters face or to the stimuli). For example, to illustrate a down turned mouth a child may state that "his mouth goes down like that" and demonstrate by pulling their own mouth down.

Nearly all children used some kind of non-verbal descriptions. The mean number of non-verbal descriptors (including and excluding repetitions) was calculated for 6, 8, and 10-year-olds and this information is shown in Figure 6.5

⁴⁷ The configural category is explored later within this section.

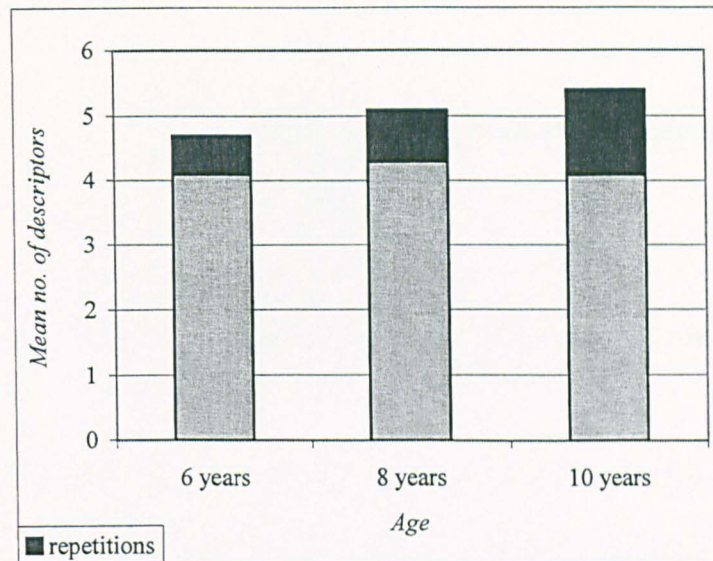


Figure 6.5: Mean number of non-verbal descriptors provided by children in Experiment 2

From Figure 6.5 it can be seen that the mean number of non-verbal descriptors increased with age. However, when non-verbal descriptors which were repetitions (of verbal data) were removed the number of non-verbal descriptors was approximately the same for children of all ages. A closer observation of the *content* of these descriptors indicated that 10-year-olds tended to provide non-verbal descriptors in addition to verbal descriptors, for example stating, “his mouth goes like that” and demonstrating the mouth going down with their own mouth, and then saying “it goes down”. Whereas 6- and 8-year-olds tended to use non-verbal descriptors in place of verbal descriptors, for example, simply stating “his mouth goes like that” and then demonstrating the mouth going down with their own mouth.

The *content* of the non-verbal descriptions (which were not repetitions of verbal data), was further analysed. Results showed that the highest single category of non-verbal descriptors for all children was for *location* (information about a particular part of a feature) for example, a child stating “his nose goes fat *there*” and pointing to the bottom

of the nose. Table 6.8 shows the percentage of non-verbal descriptors which are location descriptors.

Age	Mean % of Non-Verbal descriptors	
	Location	Other
6 years	59.57%	40.43%
8 years	50.98%	49.02%
10 years	33.33%	66.67%

Table 6.8: Mean percentage of non-verbal location and other descriptors provided by children in Experiment 2

From Table 6.8 it can be seen that that the approximately half of 6- and 8-year-olds, and a third of 10-year-olds non-verbal descriptors related to location. The remaining non-verbal descriptors consisted of children tracing or drawing the *shape* of a feature and also illustrating *distances*. Location descriptions will be considered in more detail in the following section.

Location Descriptions (Verbal and Non-Verbal)

Location descriptions were verbal (e.g. a child stating “his hair is curly on the *top*”) and non-verbal (as described above). The mean number of location descriptors (both verbal and non-verbal) for 6-, 8-, and 10-year-olds was calculated and is shown in Figure 6.6

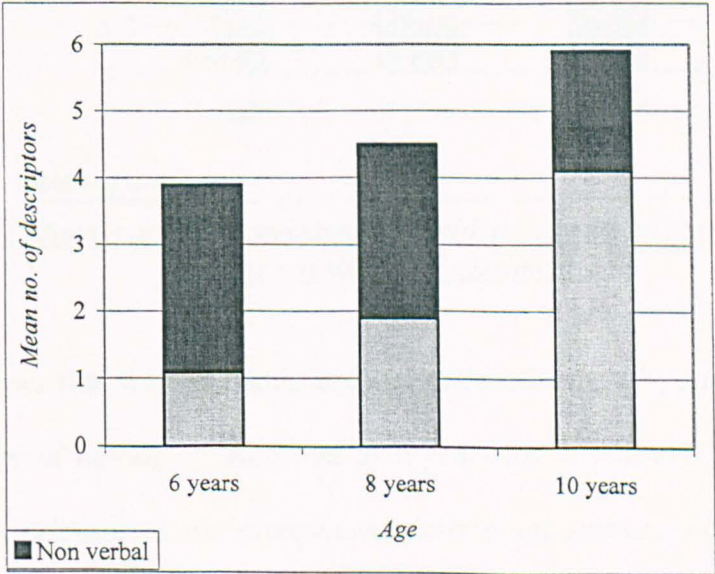


Figure 6.6: Mean number of location descriptors provided by children in Experiment 2

Figure 6.6 illustrates that on average 10-year-olds provided the highest number of location descriptors, followed by 8-year-olds, with 6-year-olds providing the lowest number of location descriptors. Whereas, the majority of 6- and 8-year-olds location descriptors were non-verbal. Less than one third of 10-year-olds location descriptions were non-verbal.

The Effect of a Comparison Task

The Effect of a Comparison Task on the Quantity of Children’s Descriptions of Unfamiliar Faces (aim (iv))

The mean number of descriptors (all terms including configural descriptions) provided by children in the comparison description section of the interview was calculated (both including and excluding repetitions). This data is shown in Figure 6.7.

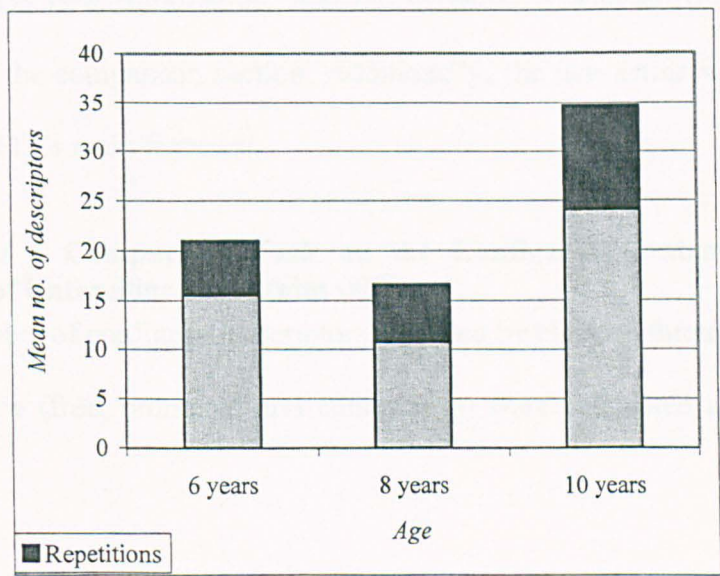


Figure 6.7: Mean number of descriptors provided by children in the comparison interview section in Experiment 2

Figure 6.7 shows that without taking account of repetitions, 10-year-olds provided the highest number of descriptors followed by 6-year-olds. 8-year-olds provided the least number of descriptors in the comparison description section. When repetitions of descriptors (usually provided in the free and prompted sections of the interview) were removed, the difference between age groups was reduced. 10-year-olds still provided

the highest number of new descriptors followed by the 6-year-olds, and 8-year-olds provided the lowest number of new descriptors (which is the same pattern described in the prompted description section of the interview).

The mean number of new descriptors used by 6, 8 and 10 year olds in the comparison description section of the interview was compared to the mean number of descriptions provided by 6, 8 and 10 year olds in the free and prompted description sections of the interview (combined), with all repetitions removed, in order to calculate the percentage increase of new information provided in the comparison description. Analyses indicate that children of all ages provided new information in the comparison section: 10-year-olds had a 53.91% increase of new information, followed by the 6-year-olds who provided 45.61% new information, and the 8-year-olds who provided 28.20% new information in the comparison section. Additionally, the new terms were provided for all seven of E-FIT's main features.

The Effect of a Comparison Task on the Configural Content of Children's Descriptions of Unfamiliar Faces (aim (v))

The mean number of configural descriptors provided by children throughout all sections of the interview (free, prompted and comparison) were calculated and are shown in Figure 6.8.

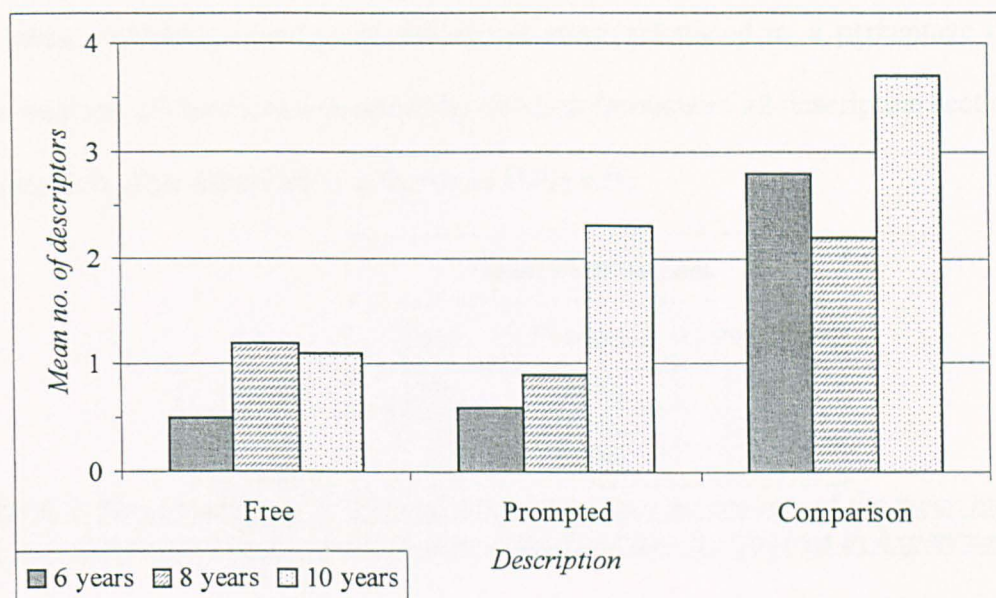


Figure 6.8: Mean number of configural descriptors provided by children in each section of the interview in Experiment 2

Figure 6.8 indicates that for all age groups the majority of configural information was provided in the comparison description section (as would be expected since this was where children were prompted for configural information). Interestingly, on average, 6-year-olds provided more configural descriptions than 8-year-olds in the comparison description section of the interview. Importantly, children from all age groups provided some configural information in the free description section (for both the primary and secondary face). 6- and 8-year-olds provided approximately the same number on average in the free and prompted description sections of the interview. In contrast, there was a difference between the free and prompted description sections of the interview for 10-year-olds.

In terms of the numbers of children providing configural information, although 6- and 8-year-olds provided approximately the same number of configural descriptors in total, only 6 children aged 6-years old provided configural descriptors whereas 8 children aged 8-years of age and 9 of the 10-year-olds provided configural descriptors.

The mean numbers of configural descriptors were calculated as a percentage of the mean numbers of descriptors provided by children throughout all description sections of the interview. This information is shown in Table 6.9.

	Interview section		
	<i>Free</i>	<i>Prompted</i>	<i>Comparison</i>
6-years	4.9%	2.5%	17.9%
8-years	6.2%	3.9%	20.4%
10-years	5.6%	9.13%	15.4%

Table 6.9: Mean number of configural descriptors as a percentage of the mean number of descriptors provided in each section of the interview by children in Experiment 2

From Table 6.9 it can be seen that configural descriptors accounted for a small percentage of children's descriptors. Although, on average, 10-year-olds provided the highest *number* of configural descriptors in the comparison section of the interview, they accounted for a smaller percentage of the descriptors provided than for 6- and 8-year-olds.

The majority of configural descriptors for 6-, 8- and 10-year-olds related to Eyes and Eyebrows (in relation to each other). The remaining configural descriptors referred to Hair Length, Mouth, Nose and Ears⁴⁸.

6.4 DISCUSSION OF STUDY 2

6.4.1 DISCUSSION OF STUDY 2 RESULTS

Analyses of the results initially looked at the descriptors children were using to describe unfamiliar faces. Analyses then went on to examine the quantity of descriptors children provided. Finally the effect of a comparison task on the content and quantity of descriptors provided was considered.

⁴⁸ The majority of non configural descriptions provided in the comparison description section related to Face Shape for 10- and 6-year-olds and Eyes for 8-year-olds.

The Language and Terms Children use to Describe Unfamiliar Faces (aim (i))

The results from Study 2 indicate that for the majority of facial features children's descriptors did not match the existing E-FIT (adult) terms and only a low percentage of E-FIT terms were used. Additionally, only a low percentage of children who provided descriptions used E-FIT terms. Similar preliminary findings have also been recently reported by Memon (A. Memon, personal communication, 14th May 2004).

When analysing the descriptors children used, a number of patterns emerged which occurred across all conditions of the interview and across all ages, suggesting a quantitative rather than a qualitative shift in performance with age. Pattern 1 referred to children frequently describing a feature element using a large variety of descriptors which matched E-FIT terms. Such a pattern avoids the problems which can occur in the communication process between an adult operator and a child witness as described in Chapter 4 and highlighted by operators' responses to the questionnaire survey, for example, when operators are required to interpret witnesses' descriptions.

For Pattern 2, children frequently described a feature element and used a small variety of descriptors which could be matched to E-FIT terms. Even though children were using the same descriptors as the E-FIT (adult) terms, they were not using them in an equivalent way. For the feature element Face Shape children used only two descriptors (round and oval) and applied them to a much broader range of different face shapes. This is consistent with previous research which suggests that children include a broader range of meanings within their use of one word, that children can have a much more restricted use of words (Jones, 2003), and that children may use words differently from adults (Saywitz, 1989; Saywitz *et al.*, 1990).

For Pattern 3, children frequently described a feature element and used their own descriptors which could not usually be matched to E-FIT terms. Children were using their own descriptors which were restricted to each individual feature element. Interestingly, in Experiment 1, many adults provided a number of subjective descriptions of composite faces for example “he looks like Colombo”. Generally, children did not include any subjective descriptions, apart from one child stating that one of the composites looked “like my tennis teacher”. The lack of subjective descriptions from children may be viewed as a positive finding as although subjective descriptions such as these may aid in face recognition, they constitute poorer verbal descriptions when used by a second party (Wells and Hryciw, 1984). Furthermore, children not using subjective descriptions may assist the translation process the operator has to go through. Finally, a number of feature elements were categorised as Pattern 4, where children did not provide any descriptors (e.g. Eyebrow Size and Eyebrow Colour).

Only one feature element was categorised as Pattern 1 (Hair Length). For the remaining feature elements which were categorised as Patterns 2 to 4, children and operators require interventions in order to enhance the communication process when constructing a composite. As described in Chapter 2, the use of some sort of prompt may augment children’s responses by increasing the number of descriptions provided by children which match E-FIT. As described above, children may comprehend terms even if they do not use them. Observations of the data provide preliminary evidence for this in that children use a number of terms for a feature element but do not use the same (applicable) terms for another feature element. For example, the feature element Hair Colour was categorised as Pattern 2 whereas the feature element Eyebrow Colour was categorised as Pattern 4, although children could have used the same terms for either

feature element. The descriptors children used may be adapted as age appropriate terms for use with children. These descriptors could be easily mapped onto the E-FIT feature index, thereby reducing the need for an operator to ‘translate’ children’s facial descriptions.

The Quantity of Children’s Descriptions of Unfamiliar Faces (aims (ii) and (iii))

When analysing the quantity of descriptors provided by children, the interviewer observed that the older children (10-year-olds) appeared to be talking a lot more than the younger children. The analyses showed that indeed, older children were providing more descriptors than younger children. However, they were also repeating more information than the younger children (10-year-olds were repeating on average over three times as much as 6-year-olds). Previous studies which have examined the quantity of children’s descriptions (e.g. Leippe *et al.*, 1991) have found statistically significant effects of age, although such studies usually include repetitions of data.

The hypothesis that older children would provide statistically significantly more descriptions than younger children was not supported due to the large variation within age groups. This is consistent with the view described in Chapter 2 that age is only one factor in the determination of children’s performance (Baker-Ward, 2002; Davies, 1996) and that individual differences regarding children’s testimony may explain within-age variation in children’s interview performance (e.g. Bruck and Ceci, 1999). A recent study by Paterson, Bull and Vrij found statistically significant relationships between individual differences in cognitive abilities of higher complexity and in social competence and the amount of correct information recalled throughout the interview in 5- to 10-year-olds (B. Paterson, personal communication, 24th May 2002).

Consistent with previous research which suggests that the completeness of free recall increases with age (see review by Davies and Westcott, 1999) 6-year-olds provided statistically significantly fewer descriptions than the older children in the free description section. Furthermore, consistent with previous research reviewed by Davies and Westcott (1999), the hypothesis that prompting would statistically significantly increase the quantity of children's descriptions was supported, suggesting the need for specific prompts with children. Although there was no statistically significant age by prompt interaction the use of specific prompts increased 6-year-olds descriptions to a level on par with that of the 8 and 10-year-olds. Therefore the use of such specific prompts appeared to be of greatest benefit for the youngest age group of children. In addition, the absence of a statistically significant age by description interaction indicates that children of all ages are capable of providing important details, regardless of age.

Although there was not a statistically significant effect of face order, on average, children provided a higher number of descriptors for the primary face than the secondary face they were shown. A possible explanation of this may be that children had become tired by the secondary face (following the practice task and the description of the primary face). The fact that children provided fewer descriptions on the second face they were shown may be used to argue against showing children more than one face in a single experimental study or in a forensic interview. However, showing children two faces in the present study enabled observations which would not otherwise have been made. For example, results indicated a statistically significant face order by description interaction. Although the number of descriptions for the second face children were shown were lower on average than for the first face, free descriptions were higher for the second face. It was the oldest age group, the 10-year-olds, whose free recall descriptions improved the most. The 10-year-olds appear to be able to utilise

the specific prompts they were provided with for the first face they were shown. Perhaps it was a result of including a higher number of free descriptors for the secondary face that children's prompted descriptions contained less information, resulting in fewer descriptors for the secondary composite face overall.

The Effect of a Comparison Task on Children's Descriptions (aims (iv) and (v))

Results showed that the comparison task proved productive for children of all ages in terms of increasing the number of descriptors, in particular configural descriptors. Although no specific hypotheses were stated in relation to this aim, based on the research cited in Section 3.2.2 it was expected that all children's descriptions would contain configural information. Indeed, children of all ages provided configural descriptions and the quantity provided increased with age. This is consistent with contemporary research that younger children process faces in a manner that is qualitatively similar to older children and adults (e.g. Gilchrist and McKone, 2003). The differences across the different age groups of children suggest a quantitative change rather than a qualitative change in the use of configural descriptions.

The majority of configural descriptors were provided in the comparison section of the interview (as would be expected as this is where children were prompted for this). Interestingly, children from all age groups provided some configural information in the free description section of the interview (for both the primary and secondary faces shown). This may be explained by the use of a practice task (where children were asked to compare two different houses in terms of configural information). In the same way children utilised the prompts provided for the primary face when providing a free description of the secondary face, they utilised the configural prompts from the practice task.

Although the composite task proved productive in terms of both the quantity and content of children's descriptions, it may not suit a forensic setting, given the potentially contaminating effect of the non target face on witness memory. The practice task may however be utilised in such a setting (Home Office, 2002).

Additional Observations

A number of children used non-verbal descriptors. Specifically, the oldest children used non-verbal descriptors in addition to their verbal descriptors, whereas the younger children (6- and 8- year-olds) used non-verbal descriptors in place of their verbal descriptions. Research has shown that the vocabulary for describing faces has frequently been described as insufficiently rich or precise for it to be an effective way to communicate information about faces (Phillips, 1978), and verbal labels do not exist because stimulus attributes do not lend themselves easily to verbal associations (Bourne, 1966). For these reasons, Garbarino and Stott (1992) suggested that "there are limits to what children under seven can convey to others using words alone" (p. 67). Even though children are provided with specific prompts, and results show they benefit from these prompts, language is still placing limits on the extent to which children can communicate their experiences. Therefore, children's use of non-verbal descriptors may be a result of their language development (and in particular their restricted vocabulary for facial descriptions) limiting their recall skills even though the information is available. The use of non-verbal descriptions may make it easier for children to communicate to an interviewer by demonstrating.

The main use of non-verbal descriptors by children was to provide details of the location of a feature description. From existing research with adults, it is known that the adjectives used to describe faces are largely size and shape words which are high frequency words. Most importantly, descriptors are largely feature independent, where

words are suitable for describing any shape such as a building and relate to size (Laughery *et al.*, 1986). Facial patterns and configurations are extremely difficult to represent in words (Clifford and Bull, 1978). Therefore, there are few words to describe location information which may prove particularly problematic for children due to their language development as noted above.

The findings from the present study highlight the need for specific prompts (especially for the youngest age group of children). The large individual differences within age groups indicate that children should not be excluded from providing facial descriptions simply on the basis of their age. An analysis of the content of children's descriptions illustrated that children of all ages use a combination of both E-FIT (adult) terms and their own language and terms (including non-verbal descriptions) to describe an unfamiliar face. Additionally, children of all ages are able to provide configural facial descriptions. Collectively, the findings from the present study demonstrate that children are able to describe unfamiliar faces. Importantly, even at age six, children are able to describe many of the key facial features, which is a necessary requirement for composite construction.

6.4.2 LIMITATIONS OF STUDY 2

A criticism of the free labelling method employed in this study is that it may underestimate children's underlying knowledge by concealing their potential ability. For example, children may understand a term but not use that term to describe a feature. In their research on facial expressions, Widen and Russell (2003) suggested that a child might know the correct emotion category but not know its label. That is, if at a certain age the correct emotion word is simply unavailable (i.e. is not in the child's vocabulary) then the free labelling task is inappropriate. This applies equally to the present study. If

a child produces a descriptor, then that descriptor is clearly available to that child. However the converse is not true. Perhaps a child for whom the term was available did not find it appropriate for the stimuli used in this study. However, as described in Section 2.1.3, children's vocabulary expands faster than their comprehension of language, and children frequently use words before they know their adult meaning. The mere fact that a word is part of a child vocabulary by no means indicates that they understand the word (Walker, 1999). Therefore, care must be taken with the application of the terms and language children used in the present study.

Non-verbal descriptions were established by notes made by the author throughout each interview in order to supplement the audio recording and this method of recording may therefore have underestimated of the number of non-verbal descriptions provided by children. A more accurate measure of non-verbal descriptions could be taken by video recording interviews. ABE (Home Office, 2002) also stresses that interviews with child witnesses should be video recorded, following this guidance could make the interview process forensically more valid.

The discussion of the results highlights the possible effect of the practice task on the content of children's descriptions, in particular children's inclusion of configural information in their free descriptions following the practice task. However, the practice task was not included as an independent variable effect (as the positive effects of practice interviews have already been established (e.g. Poole and Lamb, 1998)). Therefore possible benefits of the use of houses as practice stimuli could not be assessed here.

6.4.3 SUMMARY: IMPLICATIONS OF STUDY 2 AND FUTURE RESEARCH

A number of issues have been raised which have implications for future research. Some of these issues can be isolated and considered in terms of the current thesis and others can be highlighted as future directions for research in general, which are beyond the scope of the current thesis. The latter will be considered first.

Transcriptions of children's interviews produced a wealth of data available for extensive future analysis. For example, the feature saliency of children's descriptions may be examined in terms of free and prompted descriptions. The comparison task proved useful in promoting configural information and increasing the amount of information provided. However, such a task would not be appropriate in a forensic setting. Therefore, future research examining the effects of the comparison task and perhaps possible developments of such a task would be useful. Additionally, the development of a practice task to increase configural descriptors may be investigated. Finally, non-verbal descriptions offered by children could be examined in more detail by videoing the interview.

The main issue to be considered in terms of the current thesis is the careful development of children's facial descriptions collected in the present study as an age appropriate way of verbally prompting children in order to augment their facial descriptions, and subsequently enhancing the communication process between a child witness and an operator when constructing a composite. In addition to these verbal prompts, non-verbal prompts may also facilitate communication process by reducing the language demands of the task (Pipe *et al.*, 2002) and potentially providing a way of categorising non-verbal information. It is the development of such techniques which will be used to direct the following studies in this thesis.

CHAPTER 7: STUDY 3

THE EFFECT OF PROMPTS ON CHILDREN'S DESCRIPTIONS OF UNFAMILIAR FACES

This chapter describes Study 3, an empirical investigation of the effect of sets of visual and verbal prompts on children's descriptions of unfamiliar faces. Study 3 comprised two experiments: Experiment 3 involved adult participants' selection of the prompts for target faces which remained in view, in order to provide a measure of the 'correct' prompts for later coding of child participant responses in Experiment 4. Experiment 4 describes child participants' selection of the prompts for target faces from memory.

7.1 INTRODUCTION

The findings from Studies 1 and 2 suggested that there may be techniques which can assist both children and operators in the facial composite interview. Responses to Study 1 identified children's concentration span and language development as the most frequently reported issues faced by operators when interviewing children. Research findings cited in Chapters 2 to 4 illustrated how these issues can pose problems at both the level of the witness and of the operator. The findings from Study 2 indicated that children's free facial descriptions were particularly limited, and for the majority of facial features, children's verbal descriptors did not match the existing E-FIT (adult) terms. Children of all ages used a combination of both E-FIT terms and their own descriptors to describe unfamiliar faces. Additionally, younger children were less able to communicate verbally than older children. In order to assist the communication process between operators and child witnesses there is a need to increase the number of

descriptors provided by children which match or correspond to the E-FIT feature index and to utilise the non-verbal information provided by children.

Study 2 analyses also indicated that with appropriate prompting children (especially younger children) are able to provide facial descriptions. In response to the questionnaire survey operators reported using additional materials in order to assist witnesses in their descriptions. As described in Section 2.3.3 the use of some sort of non-verbal prop or prompt may augment children's (in particular, young children's) brief responses to open-ended questions. Props and prompts can provide concrete, external retrieval cues which may improve memory searches (e.g. Pipe *et al.*, 1993), reduce the language demands of a recall task (Pipe *et al.*, 2002), help children comprehend what adults are asking them and reduce the social and emotional demands of an interview (Westcott and Littleton, in press). However, a criticism which can be generalised from previous studies (e.g. Dent and Stephenson, 1979; Goodman *et al.*, 1991; Orbach and Lamb, 2000), is that the information elicited using recognition memory prompts is more likely to be more inaccurate than information elicited using open-ended prompts by leading children into commission errors (Pipe *et al.*, 1993) and this is of particular concern with young children (Saywitz *et al.*, 1991). However, this criticism usually refers to non-verbal props which are situationally specific: that are representations of actual props present at the event being recalled or can be associated with the alleged offence. Section 2.3.3 described how non-verbal 'prompts', which are not exact replicas of original objects present at the witnessed event, can improve recall without leading to a decrease in accuracy (Pipe *et al.*, 1993) and reduce the chance that recall will become contaminated through the provision of specific information (Ling and Blades, 2000) or over elaboration (Aldridge *et al.*, 2004).

There has been a wealth of research conducted into the use of non-verbal prompts (and indeed props) with children, including a small amount of research in the area of suspect identification (e.g. Goodman *et al.*, 1991). Limited research has also focussed on specific methods to aid children's recall of *person* features (including some facial features). In an attempt to improve children's testimony for a target's age, Rudy and Goodman (1991) and Tobey and Goodman (1992) constructed an age identification task. Children were provided with a line-up of photographs of persons representing different age ranges. Results showed that children gave more correct responses when picking a photograph than they did when answering a free recall question and the line-up was particularly useful for the very young children. When simply asked the man's age, 88% of the 4-year-olds said "don't know" whereas 70% of them could point to the correct picture. However, children who gave no response or said "don't know" in response to the free recall question often chose an incorrect photograph in the recognition task.

Schwartz-Kenney, Bottoms and Goodman (1996) extended the above research to include other person features and some facial features. They developed a series of computer generated 'feature line-ups' consisting of a series of pictures printed on cards containing separate representations of typical instances of features. Features included Age, Eye Colour, Build, Hair Colour, Hair Texture and Skin Colour. The Age and Skin Colour line-up consisted of magazine photographs of women. Eye Colour, Build, Hair Texture and Hair Length line-ups were represented by simplistic line drawings. The Hair Colour line-up consisted of five swatches of synthetic hair. Finally, the height line-up was composed of lines placed on a wall (similar to the technique found to be useful by Saywitz *et al.*, 1991). Additionally, there were three versions of the Hair Length line-ups, one for each of the three hair texture types. Thus, if a child initially picked wavy

Hair when asked about Hair Texture, he or she would only be shown the wavy set of length drawings when asked about Hair Length.

Schwartz-Kenney, Bottoms and Goodman conducted a first experiment where children participated in interactive play with an unfamiliar target. When questioned about the appearance of the target, results showed that children provided statistically significantly more correct responses when using line-ups than when simply asked a free recall question. Older children (8- to 9-year-olds) provided more correct responses than younger children (5- to 6-year-olds), although these differences were only statistically significant for certain features (e.g. Age and Height line-ups were more useful for the older children than the younger children). Schwartz-Kenney, Bottoms and Goodman noted that the use of magazine pictures (for the Age and Skin Colour line-ups) contained additional differences than the feature in question and therefore may have distracted the children, thereby suggesting that the task needed to be further simplified. For example, the Age line-up consisted of three women of three distinct ages but also with three different hair colours, hair textures, and so on.

Schwartz-Kenney, Bottoms and Goodman also noted that the line-ups statistically significantly decreased the children's overall "don't know" responses compared to free recall. However, in some cases the increase in responses led to an increase in commission errors: children gave significantly more incorrect answers overall in response to the line-up task as compared to the recall task. Schwartz-Kenney, Bottoms and Goodman found that the increase in commission errors was only significant for the Age, Height, Eye Colour, and Weight line-ups. The increase in commission errors is consistent with the previous research usually related to non-verbal *props* (described in Section 2.3.3). Additionally, Rudy and Goodman (1991) and Tobey and Goodman

(1992) also noted that when children responded with silence or a “don't know” response during recall, they often would then choose an incorrect photo when using a line-ups. Research by Dekle, Beal, Elliott, and Huneycutt (1996) reported that children who were presented with a target-present line-up were less likely than adults to use a “don't know” option. Dekle *et al.* and Schwartz-Kenney, Bottoms and Goodman claimed that these findings are due to the demand characteristics of the tasks in that children are not overtly presented with the physical option of responding “I don't know.”

In a second experiment Schwartz-Kenney, Bottoms and Goodman included a card with a question mark as an explicit option to represent “I don't know” in every line-up. Children were trained to choose the card with a ‘?’ on it to indicate if they did not know the answer. Results indicated that the use of the ‘?’ card led to fewer inaccurate responses. The frequency of commission errors on the line-up task compared to that on the free recall task increased only for Height and actually decreased for Hair Texture. In their discussion, Schwartz-Kenney, Bottoms and Goodman (1996) stated that children were comfortable with selecting the card with the question mark on it to represent “I don't know” when it was presented as an option among the other line-up items.

Finally, the early research on the development of face processing described in Section 3.2.2 indicated that younger children's face perception uses a predominately featural strategy (Carey and Diamond, 1977; Diamond and Carey, 1977). On the basis of this early research, Schwartz-Kenney, Bottoms and Goodman predicted that it would be beneficial for children to begin the person identification process by recognising individual features. However as illustrated in Chapter 3, contemporary research indicates that young children are perceiving and remembering faces holistically (e.g. Donnelly and Hadwin, 2003; Gilchrist and McKone, 2003). Indeed, Schwartz-Kenney,

Bottoms and Goodman concluded that their findings did not support the assumption that children process person information in a featural way as Diamond and Carey reported. They went on to suggest that perhaps children should report information about a stranger as a whole rather than in parts. Consistent with this statement, research by Turner (2004) provided some evidence that constructing composites using a featural strategy may not produce composites which are as accurate as when a holistic strategy is used. Turner found that the use of a schematic minimal face during the composite construction process reduced any disadvantage of using a featural strategy. The premise of Turner's work was that the presence of a face-context would invoke the face-processing system to a degree that an individual feature would not, thereby facilitating composite construction. The schematic minimal face used by Turner consists of a simple outline with two circles for Eyes and short lines for the Nose and Mouth. The shape of the outline and the size and position of the features were designed to reflect the average shape, size and position of the corresponding features within E-FIT as closely as possible⁴⁹.

There is a need to extend the limited research described regarding facial descriptions and facial composite research. If children's retrieval of information about past events and person descriptions can be enhanced with prompts, their retrieval of facial descriptions may also be improved with a recognition technique specifically related to facial features. Schwartz-Kenney, Bottoms and Goodman viewed their line-ups as an alternative to the recall tasks of composite systems: they were used in isolation and they did not map onto any composite systems. However, further research may investigate the use of such visual prompts in addition to a composite task. For example, prompts may be used to obtain initial facial descriptions which map directly onto a composite system,

⁴⁹ The minimal face is currently available as part of Aspley E-FIT (version E-FIT.NET).

thereby further reducing the problems of language interpretation and translation between a witness and an operator.

7.1.1 EXISTING PROMPTS USED BY COMPOSITE OPERATORS

In response to the questionnaire survey in Study 1, four police forces attached the forms they use to collect witnesses descriptions. These forms included verbal prompts which consisted of one to two word open prompts relating to various facial features which broadly correspond to the E-FIT description boxes. These prompts sometimes included a selection of alternative descriptors to assist the witness and police officer. In addition some forms included visual prompts. Personal correspondence with E-FIT operators (A. Parry and C. Charsley, 2nd November 2001; J. Richardson, 14th October 2002) also revealed that they used a number of visual prompts used when interviewing witnesses. The available prompts are described below:

E-FIT

E-FIT's feature index contains a number of illustrations to assist witnesses of their descriptions of features which "cannot be adequately expressed in words" (E-FIT v3.1 for Windows User Manual, p.59). These include the length and width of Face (combined), Lip thickness and Mouth (combined), Ears and Noses. However, no operators reported using these illustrations in Study 1.

Study 1 indicated that approximately half of the operators reported showing E-FIT's description boxes to adult and child witnesses in an attempt to prompt the witness for a more detailed description and to focus and involve witnesses. Clark, Pike, Brace and Kemp (2001) allowed half of their adult participant witnesses to view E-FIT's description boxes, and if they wished they could add to or modify their descriptions having seen them. It was stressed that they did not have to choose an option in every

characteristic box and could leave as many as they wanted blank. Findings demonstrated that allowing witnesses to view the description boxes did not have a detrimental effect on the quality of their composites, as long as they were provided with the option of a “don’t know” or “can’t remember” response (i.e. leaving a blank box).

Metropolitan Police

During an initial CI, Metropolitan police cover the majority of E-FIT’s description boxes using a scale (usually of 1 to 5) for sizes and shapes of features. The use of this scale is followed by a number of visual prompts for use with witnesses. These visual prompts consist of line drawings of Face Shape, Fleshiness of Face, Eyebrow Thickness, Eyebrow Spacing, Eyebrow Shape, ‘Openess’ of Eye (covers Eye Size and Eye Shape), Spacing of Eyes, Pupils, Nose Bridge and Nose Tip, Mouth Shape, Lip Thickness, Chin Shape, Facial Hair, Profile and Teeth. Additionally, the Metropolitan police use a drawing of a blank face on which the witness or the operator can draw facial features (e.g. Hair, Hairline, Scars) and a number of bars of varying shades of grey upon which the operator can mark the Skin and Hair tones. Figure 7.1 shows an example of the visual prompts for Mouth Shape shown to witnesses by the Metropolitan Police E-FIT operators.

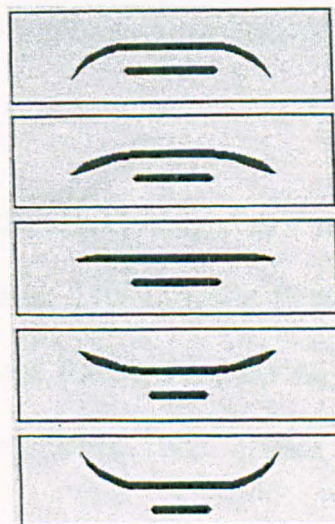


Figure 7.1: Example of Mouth Shape prompts from the Metropolitan police

Federal Bureau of Investigation Catalogue

The Federal Bureau of Investigation (FBI) in the USA publishes a Facial Identification Catalogue for composite artists. This Catalogue is available to sketch artists in the UK to assist witnesses in their descriptions, and may also be used for the same reason by E-FIT operators (J. Richardson, personal communication, 14th October 2002). The catalogue is a large reference source of photographs of 13 main features, including Head, Eyes, Eyebrows, Nose, Mouth, Chin, Cheek, Ears, Hair, Facial lines, Scars, Forehead, and Skin Irregularities. Within these main feature categories there are 60 features elements. Each photograph is partially obscured in order to focus a witness's attention on the feature element in question. Figure 7.2 illustrates an example of the visual prompts for Nose in the FBI catalogue.

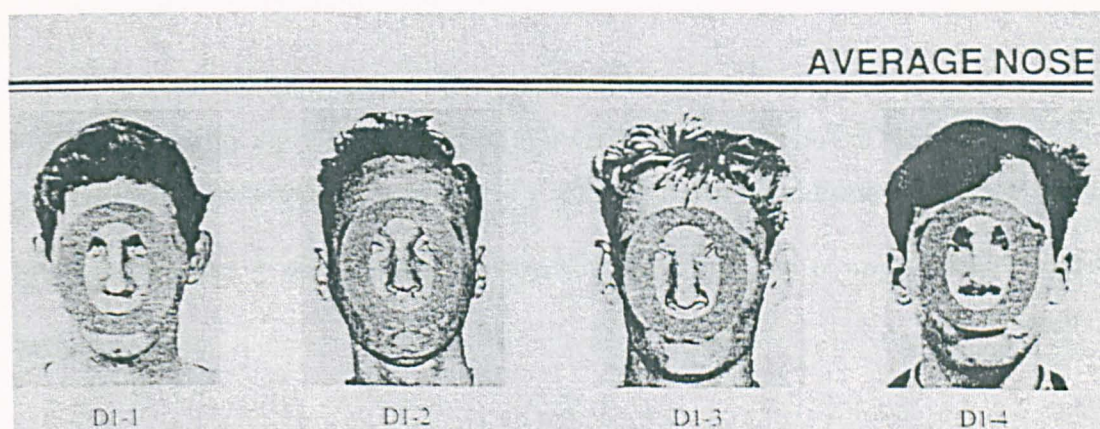


Figure 7.2: Example of Average Noses from the FBI catalogue

Identikit Catalogue

Some sketch artists also use the Identikit Catalogue as a source of visual prompts (J. Richardson, personal communication, 14th October 2002). The identikit catalogue provides over 100 pages of isolated feature selection for Hair Style, Chin Shapes, Eyebrows, Eyes, Nose, Lip Shapes, Facial Tone, glasses, facial hair, age lines and

headwear. Each page consists of between four and 12 sketched images. Figure 7.3 shows an example of the visual prompts of Identikit used with witnesses.

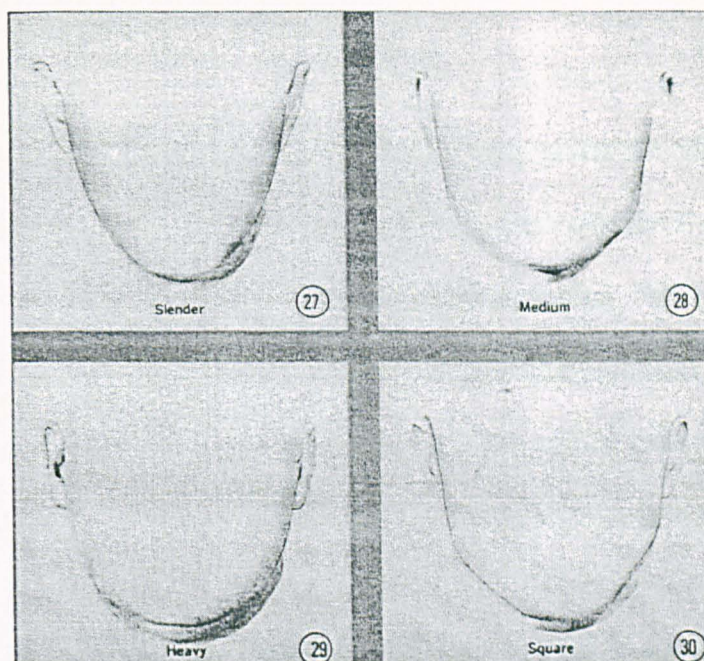


Figure 7.3: Example of Chin prompts from the Identikit catalogue

7.1.2 SUMMARY OF EXISTING PROMPTS AND ASSOCIATED ISSUES

As described in the discussion of Study 1, various forms of additional materials are used by operators and there is a lack of standardised format or prompts used within and between forces.

The use of the five-point scale (depicting size and shape) by the Metropolitan police can be confusing for child witnesses (A. Parry and C. Charsley, personal communication, 14th November, 2001). A personal experience of an E-FIT interview using such a scale also proved confusing for the author, due to the need to visualise the scale and make choices before the visual prompts were shown (which did not map on to the five-point scale). The use of both a scale and visual prompts requires a witness to repeat descriptions in a different way. For example, a witness may be asked the size of the suspect's eyes on a scale of 1 to 5. They will later be shown visual prompts illustrating

only three different sizes of eye. Children may assume that any previous answer they gave was incorrect and may alter their response when they are asked the same questions more than once (Siegal, 1991).

Although gaining detailed information during the CI may assist the witness in their recall, only a small amount of a witness's description is used with the E-FIT system in order to create a composite (A. Parry, personal communication, 7th January 2003). When considering the visual prompts, a number of them are not used by the E-FIT program (for example the Pupil and Profile prompts). The use of such prompts which will not be entered into E-FIT's description boxes results in redundant information.

Finally, none of the visual prompts described map onto the E-FIT's description boxes which an operator has to subsequently use, resulting in a further reliance on operators to interpret witnesses' descriptions (as described in Chapter 4 and illustrated by Figure 4.2)⁵⁰. For example, the visual prompts for mouth shapes used by the Metropolitan police consist of five drawings: down turned mouths (2 drawings); straight mouth (1 drawing); upturned mouths (2 drawings) (as illustrated in Figure 7.2). However, in the E-FIT description boxes there are 5 choices for mouth shape which are Upturned, Straight, Down Turned, Misshapen and Lopsided. All of these points will impact on the time spent obtaining a description from a witness, which was one of the main issues when interviewing children raised by operators in response to Study 1.

The visual prompts in the FBI catalogue are photographs and therefore are very explicit which may cause problems of suggestion (as described in Section 2.3.3) with witnesses (Aldridge *et al.*, 2004). Additionally, the sheer volume of choices in the FBI catalogue

⁵⁰ Although sketch artists do not need to translate the witnesses choice of prompts into an existing database, many sketch artists are now using the materials described with E-FIT and are having to interpret witnesses descriptions.

may cause problems: witnesses can be daunted by a high number of choices of features (J. Richardson, personal communication, 14th October 2002).

Additionally, whereas the prompts used by the Metropolitan police and the features in Identikit Handbook are all shown in isolation rather than within a whole face context, the FBI prompts are shown within a whole face. However, although all of the features apart from the feature the witness is focussing on are obscured, a witness can also see other features on the face, which may be confusing (as suggested by Schwartz-Kenney, Bottoms and Goodman, 1996). Some witnesses feel more comfortable using the FBI Catalogue but there are also occasions when a witness may get confused looking at an image within the 'wrong' face or with associated features and prefer to look at isolated features. Some witnesses are happy to use both kinds of prompts, it seems an individual preference but operators find it very useful to offer witnesses a choice (J. Richardson, personal communication, 14th October 2002). Research by Tanaka and Farah (1993) demonstrated that adults were more accurate at identifying parts of faces presented within a whole face than they were at identifying the same parts presented in isolation. Additionally, Maurer *et al.* (2002) showed that adult participants were approximately 10% more accurate in recognising the identity of a feature when it was presented in the context of the whole face rather than as an isolated feature. As described previously, contemporary research indicates that young children are perceiving faces holistically and therefore these findings may be equally applicable to them.

In summary, a variety of visual prompts are currently used by operators when interviewing witnesses in order to construct facial composites. However, a number of issues have been raised in relation to these prompts. These issues, taken together with the findings from Studies 1 and 2 emphasise the need for the development of visual

prompts which will consider and address these issues and will enhance the communication process between a witness and an operator.

7.1.3 RESEARCH AIMS OF STUDY 3

The aim of Study 3 was to develop age appropriate recognition prompts for obtaining basic facial descriptions of unfamiliar faces for subsequent use with the composite system E-FIT. A set of visual prompts and a set of verbal prompts were designed in order to assist (i) witnesses - in providing facial descriptions and interpreting operators' prompts and (ii) operators - in eliciting and interpreting witnesses' descriptions. In particular, the visual prompts were designed to provide a witness with a choice of responses which were not open to interpretation in the same way that verbal prompts are. Therefore enabling operators to be clear exactly 'how' a witness remembers a feature, rather than how they describe the feature.

(1) The main aim of study 3 was to investigate the effect of a set of prompts, and in particular to explore any differences between using a set of visual prompts or a set of verbal prompts on the:

(i) quantity of children's descriptions of unfamiliar faces. It was hypothesised that using prompts would lead to a statistically significant increase in the quantity of children descriptions.

(ii) content (in terms of accuracy) of children's descriptions of unfamiliar faces.

This aim was treated as exploratory and no hypotheses were stated.

(iii) time taken to obtain the prompted descriptions. This aim was treated as exploratory and no hypotheses were stated.

(2) Additional aims of Study 3 were to investigate:

(iv) *children's use of the "don't know" prompts.* This aim was treated as exploratory and no hypotheses were stated.

(v) *children's performance at an identification task involving the target described.* This aim was treated as exploratory and no hypotheses were stated.

For all of the aims the effect of different age groups was explored.

7.1.4 THE DESIGN AND ADMINISTRATION OF VISUAL AND VERBAL PROMPTS

On the basis of the collation of materials described in Section 7.1.1, along with the data from Study 2, prompts were designed for E-FIT's seven main features and 20 of E-FIT's feature elements. Four prompts were also produced for the configuration of features in terms of second order relations (representations of the spatial relations among individual facial features)⁵¹ (Maurer *et al.*, 2002; Pellicano and Rhodes, 2003). The visual and verbal prompts were designed and used with a number of factors in mind. Prompts were:

- designed as *representations* in order to avoid the problems associated with them being too explicit (Aldridge *et al.*, 2004) and to reduce the chance of recall which is contaminated through the provision of specific information and over elaboration (Ling and Blades, 2000)
- option-posing rather than forced choice by explicitly presenting children with the option of responding "don't know" (Clark *et al.*, 2001). This option was represented by a question mark for the visual prompt (Schwartz-Kenney, Bottoms and Goodman, 1996), and the words "not sure" for the verbal prompts. In addition, children were provided with a short training program in the importance of saying "I

⁵¹ For use at a later stage in the E-FIT program.

- don't know" (Saywitz *et al.*, 1993; Home Office, 2002) and practiced the use of the "don't know" option (Schwartz-Kenney, Bottoms and Goodman, 1996).
- designed to map directly onto the E-FIT composite system, thereby reducing the reliance of an operator to 'translate' children's descriptions in an attempt to further reduce the problems of language and time highlighted by operators' responses in Study 1 and the problems with child witnesses described in the earlier chapters. The colour feature elements (Hair Colour, Eyebrow Colour and Eye Colour) were the exception to this. For these elements a colour spectrum was used. The colours selected had to be mapped onto the E-FIT descriptions of light and dark for colour (which would have to be carried out by E-FIT operators)⁵².
 - designed to be introduced following free descriptions in order to minimise the risk of contamination and to ensure the forensic value of the information elicited earlier in the interview was not compromised (Aldridge *et al.*, 2004).
 - presented one feature element at a time, with the options for each feature element being presented simultaneously. This allowed children to clearly focus on one feature at a time in order to control for any differences which may have distracted children from concentrating on the feature in question (Schwartz-Kenney, Bottoms and Goodman, 1996).
 - numbered, printed on A4 white paper, laminated and presented within an A4 folder within which they could be easily worked through.

Visual Prompts

In order to create prompts which were visual *representations*, the schematic minimalist face (Turner, 2004) was used to present features and therefore avoid the problems

⁵² The alternative to this would be for children to change the colour they reported into grey scale which may be problematic.

associated with prompts which are too explicit (Aldridge *et al.*, 2004). All visual prompts were:

- presented within a whole face context (Maurer *et al.*, 2002; Tanaka and Farah, 1993), with the exception of the colour and Hair Type prompts which were presented in isolation as any alteration does not have an effect on the overall configuration of the face.
- presented to children with the shape feature element first followed by the size feature element (A. Parry, personal communication, 7th January 2003). Therefore, the following size feature element related to the shape exemplar children selected (Schwartz-Kenney, Bottoms and Goodman, 1996). For example, if a child selected 'arched' Eyebrows the prompts for Eyebrow size (thin, average, thick) would be presented as arched eyebrows.
- grey scale, with the exception of Hair Colour; Eyebrow Colour and Eye Colour. For prompts presented within a minimal face, the feature the witness was focusing on was black, while the remainder of the features were grey.

An example of the visual prompts for Face Shape can be seen in Figure 7.4⁵³. The complete set of visual prompts is shown in Appendix IVa.

⁵³ The verbal prompts for Face Shape are illustrated in Figure 7.5.

7.1 Face Shape

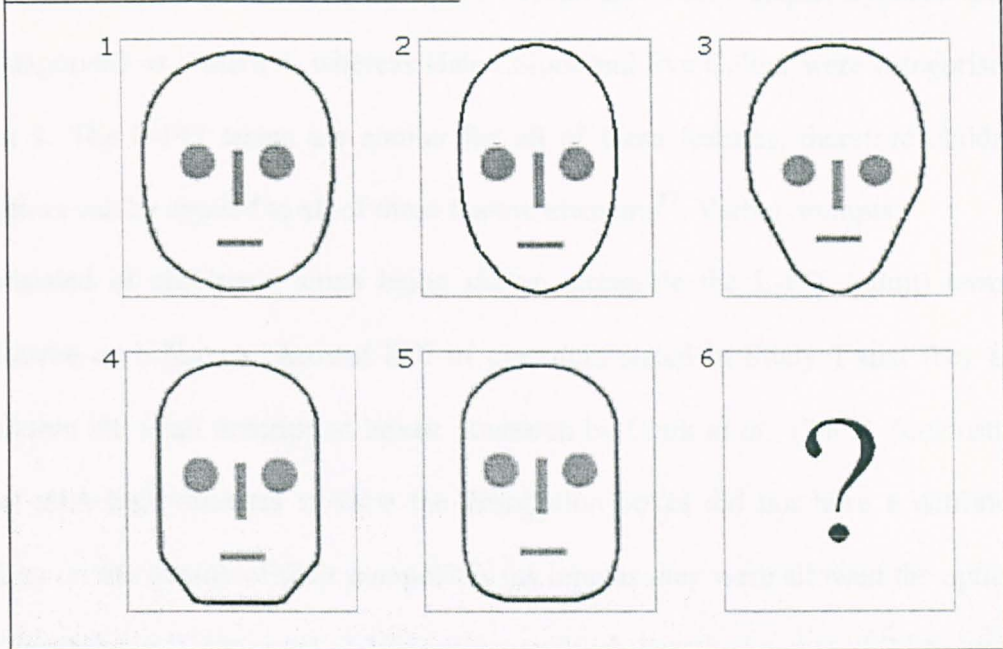


Figure 7.4: Visual prompts for the E-FIT feature element of Face Shape for use in Study 3

Verbal Prompts

In order to ensure that the verbal prompts were not too suggestive, an original set of age appropriate verbal prompts were designed by utilising the children's descriptions obtained in Study 2. "Prompting witnesses with response alternatives that have been shown to be used frequently in the past and that convey important distinctive information might be superior to simply rely upon spontaneously provided descriptions based on free recall" (Sporer, 1996, p.75). The terms children used to describe feature elements categorised as Pattern 2⁵⁴ or 3⁵⁵ were used directly (full details of each pattern was described in Section 6.3.2). For feature elements which were categorised as Pattern 4⁵⁶ the descriptors children used to describe feature elements categorised as Pattern 2 or 3 were adapted in the following way: a check was performed to see if children had used

⁵⁴ Children frequently describing a feature element and using a small variety of descriptors which could be matched to facial feature index.

⁵⁵ Children frequently described a feature element and using their own descriptors which could not usually be matched to E-FIT terms.

⁵⁶ Children not usually describing a feature.

the descriptors for other feature elements elsewhere in their descriptions thereby indicating that the language is in the child's vocabulary. For example, Eyebrow Colour was categorised as Pattern 4, whereas Hair Colour and Eye Colour were categorised as Pattern 2. The E-FIT terms are similar for all of these features, therefore children's descriptors can be applied to all of these feature elements⁵⁷. Verbal prompts:

- consisted of children's terms being shown alongside the E-FIT (adult) terms to describe each feature. Around half of operators stated in Study 1 that they show children the adult description boxes. Research by Clark *et al.*, (2001) demonstrated that allowing witnesses to view the description boxes did not have a detrimental effect on the quality of their composites (as long as they were allowed the option of a "don't know" response). Additionally, research has shown that children usually understand more words than they use; they may understand a word but not use that word to describe a feature (Widen and Russell, 2003).
- were typed (font Verdana, black, size 18)

An example of the verbal prompts for face shape can be seen in Figure 7.5⁵⁸. The complete set of verbal prompts are shown in Appendix IVb.

<u>7.1 Face Shape</u>	1. Round
	2. Oval
	3. Triangular, pointy
	4. Angular, like a diamond
	5. Square, straight
	6. Not sure?

Figure 7.5: Verbal prompts for the E-FIT feature element of Face Shape for use in Study 3

⁵⁷ Children using descriptors which could be matched to E-FIT terms. This pattern only occurred for the feature element Hair Length.

⁵⁸ The visual prompts for Face Shape were illustrated in Figure 7.4.

7.2 EXPERIMENT 3: ADULT PARTICIPANTS' SELECTION OF PROMPTS

7.2.1 METHOD

Design

A within participants design was used with one factor, 'Prompt' which had two levels (Visual and Verbal). Adult participants selected both visual and verbal prompts for two target faces. The dependent variable was the adults' selection of the prompts.

A number of measures were employed to control for order effects. The order of prompts (verbal prompts followed by visual prompts, and visual prompts followed by verbal prompts) for each target face, and the order of target face were counterbalanced across participants, with constraints such that the order and targets appeared equally frequently.

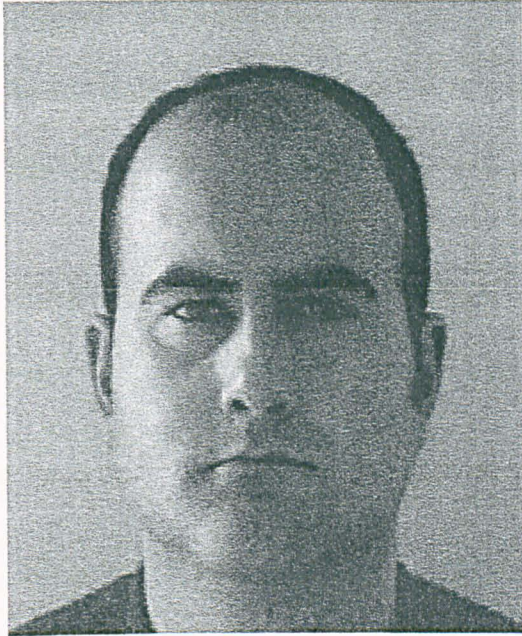
Participants

Thirty-eight adults (17 males and 21 females) were drawn from students, staff and visitors to The Open University. They were recruited informally and were not paid for their participation. Their ages ranged from 22-years to 60-years (mean=35-years, 6-months). All participants spoke English as their first language.

Materials

Two, one minute video sequences were filmed using a Canon digital video camera and edited using Ulead VideoStudio 7 for later use with in Experiment 4 with child participants. Each sequence featured a Caucasian adult male, aged between late twenties and early thirties who had consented to a video being taken and used in this study. Each target was filmed sitting at a desk and working on a computer against a plain background. The sequences featured two close-up full face poses of (15 seconds and 10 seconds).

An image of each target face was captured from each one-minute video sequence using Windows Media Player v9 and Jasc PaintShop Pro 6.0 and printed on an A4 sheet as a portrait colour print measuring 12cm by 15cm. These images can be seen in Figure 7.6a and 7.6b.



*Figure 7.6a: Target Face 1 for
Experiment's 3 and 4*



*Figure 7.6b: Target Face 2 for
Experiment's 3 and 4*

In addition to the facial stimuli, the visual and verbal prompts were used. Response sheets were also created to record participants responses. Each response sheet provided space to record adult participants' selection of prompt along with any notes for 20 of E-FIT's feature elements and 4 configural prompts. The full response sheet can be seen in Appendix IVc.

Procedure

All adults were interviewed individually. Adult participants were informed about the purpose of the study and consented. They viewed photographs of both target faces, presented individually. With each target picture remaining in view, adult participants were asked to select from both sets of prompts. The order in which the prompts were

presented (visual followed by verbal or verbal followed by visual) depended on which interview condition each participant was assigned to.

For verbal prompts, participants were asked to “select the most appropriate *word* which describes the man’s <insert the name of the feature element>” for all 20 feature elements and 4 configural selections. For the visual prompts participants were asked to “select the most appropriate *picture* which describes the man’s <insert the name of the feature element>” for all 20 feature elements and 4 configural selections.

If a participant selected the question mark or the “not sure” response for any shape feature element then the average feature was selected and used for the presentation of the following size feature element.

Following their selection of prompts, adult participants were informed about the existing procedures used by police composite operators with adult and child witnesses. They were then asked whether they would have a preference for the prompts (visual, verbal, both or neither).

All responses were recorded on the response sheets. No time limit was set, and interviews lasted between 15 and 20 minutes. At the end of the task all participants were thanked and de-briefed.

7.2.2 RESULTS

This results section considers adult participants’ responses solely to provide a measure of accuracy for the coding of child participants’ responses in Experiment 4. Additional results from adult participants will be considered in comparison to child participants’ data in the results section for Experiment 4.

Adult Participants' Selection of Prompts as a Measure of Accuracy

Adult participants' selection of the visual and verbal prompts could be used in a number of ways in order to assess the accuracy of child participants' data in the subsequent experiment. Initially adult participants' *modal* choice of visual and verbal prompt for each feature element was identified as the 'correct' selection of prompt to compare with children's selections. However, analyses of the adults' data indicated that even with the target photograph remaining in view, there were large variations between adults' selection of the prompts. Given these variations in the adults' selections a number of alternative ways to measure accuracy were considered and two measures of accuracy were explored in more detail. First, the visual comparison of the visual prompts with the target photographs by three independent judges and secondly, a method which took account of the variability of adults' responses and the possible number of selections available to a child participant. The latter of these measures was eventually selected as the most sensitive measure of accuracy and is described below⁵⁹.

Selected Method of Measuring Accuracy

The selected method of measuring accuracy involved the following technique of awarding marks to child participants:

$$\frac{\text{Actual number of adults making choice}}{\text{Expected number of adults making choice}}$$

Where the *Expected* number of adults making choice equals:

$$\frac{\text{Total number of adults } (n=38)}{\text{Total number of choices for each feature}}$$

⁵⁹ Some data from the following experiment were included as examples in Tables 7.1 to 7.5 in order to illustrate the selected measure.

Due to the fact that feature elements in E-FIT have a varying number of choices (between 2 and 7) this method corrected for the number of possible choices a participant can make. For example, the feature element Ear Shape has a total of 2 choices a participant may select (rounded; pointed). Therefore, Expected number of adults making each choice = $38/2 = 19$. Table 7.1 shows a summary of the marks that could be awarded to child participants for the feature element Ear Shape:

	Choice 1 Rounded	Choice 2 Pointed
<i>Actual</i> Number of adults making choice	31	3
<i>Expected</i> Number of adults making choice (<i>Actual/N</i>)	19 (38/2)	19 (38/2)
Mark awarded (<i>Actual/Expected</i>)	1.823529 (31/19)	0.176471 (3/19)

Table 7.1: Example summary of marks awarded to children for the feature element Ear Shape in Experiment 4⁶⁰

On the basis of this marking scheme a child who selected the prompt choice 1 ('rounded' Ears) would be awarded 1.823529 points, a child who selected the prompt choice 2 ('pointed' Ears) would be awarded 0.176471 points. Therefore, the higher number of adults who selected a choice, the greater the score that would be awarded to a child who also selected same choice.

In contrast, the feature element Hair Length has a total of 7 choices a participant may select (Bald; Almost Bald; Receding; Very Short; Short; Collar Length; Long). Therefore, Expected number of adults making each choice = $38/7 = 5.43$. Table 7.2 shows a partial summary of the marks awarded to child participants for the feature element Hair Length:

⁶⁰ Example taken from participants' selection of visual prompts for target 1.

	Choice 1 Bald	Choice 2 Almost Bald	Choice 3 Receding	Choice 4 Very Short
<i>Actual</i> Number of adults making choice	0	32	6	0
<i>Expected</i> Number of adults making choice (<i>Actual/N</i>)	5.43 (38/7)	5.43 (38/7)	5.43 (38/7)	5.43 (38/7)
Mark awarded (<i>Actual/Expected</i>)	0 (0/5.43)	5.894737 (32/5.43)	1.105263 (6/5.43)	0 (0/5.43)

Table 7.2: Example summary of marks awarded to children for the feature element Hair Length in Experiment 4⁶¹

The marking scheme was calculated for each feature element by target and by prompt (visual and verbal). Following these calculations, a maximum possible score was calculated by totalling the highest score for each feature element. This was calculated for the each set of prompts and for each target. Table 7.3 shows this information.

Target	Prompt	Maximum Score Possible
1	Visual	59.43
	Verbal	59.98
2	Visual	59.49
	Verbal	64.27

Table 7.3: The maximum score which could be awarded to child participants for each target and each set of prompts in Experiment 4

Such a calculation enabled the total mark awarded to a child participant to be worked out as a percentage of the maximum possible score. For example if a child scored 37.99 when selecting the visual prompts for target 1 this would be expressed as 63.92% ((37.99/59.43) x 100). Therefore, this allowed comparisons to be made across the two target faces and across the set of prompts (visual and verbal).

Adult Participants' Selection of Prompts as a Measure of Inaccuracy

In addition to accuracy, it is important to consider the inaccuracy of data. Adult participants' selection of the visual and verbal prompts was also used in order to assess

⁶¹ Example taken from participants' selection of visual prompts for target 1.

the inaccuracy of child participants' data. This involved the following method: one mark would be awarded to a child participant if they selected a prompt which either no adult or only one adult participant had selected. Following these calculations, a maximum possible incorrect score was calculated by totalling the number of feature elements for which there were examples of either one or no adults selecting an exemplar. This was calculated for the each set of prompts and for each target. Table 7.4 shows this information.

Target	Prompt	Maximum Score Possible
1	Visual	18
	Verbal	19
2	Visual	19
	Verbal	20

Table 7.4: The maximum inaccuracy score which could be awarded to child participants for each target and each set of prompts in Experiment 4

This calculation enabled an inaccuracy mark to be awarded to each child participant as a percentage of the maximum possible inaccuracy score. For example, if a child scored 3 when selecting the visual prompts for target 1 this would be expressed as 16.67% $((3/18) \times 100)$. This allowed comparisons to be made across the two target faces and across the set of prompts (visual and verbal).

7.3 EXPERIMENT 4: CHILD PARTICIPANTS' SELECTION OF PROMPTS

7.3.1 METHOD

Design

A mixed factorial design was used. 'Age' was a between participants factors, and had three levels (6-, 8- and 10-years), and 'Description' was a within participants factor and had three levels (Free, Visual Prompt and Verbal Prompt). All participants provided a free recall description followed by two prompt interview conditions for one of the targets. The order of interview (free description followed by prompt interview

conditions) was fixed to prevent contamination of a free description by subsequent prompts. The dependent variables were children's selection of the prompts; the time taken by children to select the prompts and children's performance at an identification task.

A number of measures were employed to control for order effects. The order of prompts (verbal prompts followed by visual prompts, and visual prompts followed by verbal prompts) and the target was counterbalanced across participants, and within age group with constraints such that the order and targets appeared equally frequently in each condition and each age group.

Participants

Ninety children (42 boys and 48 girls) were drawn from a combined school in Buckinghamshire (UK). The children were equally divided into three age groups. The first group comprised 14 boys and 16 girls from Year 1/2, aged between 5-years 7-months and 6-years 10-months (mean=6-years 2-months). The second group comprised 15 boys and 15 girls from Year 3/4, aged between 7-years 6-months and 8-years 5-months (mean=8-years 0-months). The third group comprised 13 boys and 17 girls from Year 5/6, aged between 9-years 7-months and 10-years 10-months (mean=10-years 1-month). All participants spoke English as their first language. Parental and participant consent was gained for each child participant and children were not paid for their participation.

Materials

In addition to the facial stimuli videos (described in Experiment 3) and the visual and verbal prompts, the following materials were used:

A nine-person simultaneous line-up was constructed for each of the two target faces (one target face along with eight distractor faces). Portrait colour prints (7cm by 5.5cm) of the target and other Caucasian males who resembled the target were presented on an A4 sheet of paper numbered 1 to 9 in three rows. All portraits were full face pose with a neutral expression and a blank background. Each target appeared in the same position in the array. There was no target absent array as the aim of the identification task was to determine whether or not children could recognise the target, regardless of their performance with the prompts. Each line-up can be seen in Appendix IVd.

An audio recorder with a built in microphone (Sony TCM-4ODV) was used to record children's responses. A response sheet (based upon the response sheet used with adult participants, Appendix IVc) was created in order to record child participants' selection of prompts along with any notes for 20 of E-FIT's feature elements and 4 configural prompts. An instruction sheet was also used to ensure consistent instructions given to children, and to ensure best practice guidelines for interviewing children (Home Office, 2002) were followed as far as possible.

Procedure

Pilot studies were conducted with a 7-year-old girl and a 6-year-old boy to test: the stimuli (the video, the visual and verbal prompts, and the photographic array); the practice interview; the instructions; and the feasibility of the procedure as a whole. On the basis of the pilot studies some of the instructions were re-worded to provide more details, and some additional instructions were added. For example, for the first few sets of visual prompts, each individual prompt was explained to the child in order to match the explanation of the verbal prompts, which were read out individually to the child. Pilot data are not included in the results.

In order to add to the ecological validity of the study, the study comprised a number of phases carried out over two days (as to resemble those in a criminal investigation).

Day 1

Child participants viewed a video sequence of one target in groups of up to four children. The video was played on a television in a quiet room outside of the children's classroom. Children were instructed to pay close attention to the video as they would be asked questions about it later.

Day 2

All the children were tested individually, in a quiet area of the school free from distractions. The general procedure followed the instruction sheet and is summarised as follows:

1. Introduction

Each child was introduced to the interviewer and the purpose of the study was explained. Children were informed that at the end of the interview they would return to their classroom and that, if at any time they wanted to go back to their classroom they could tell the interviewer. During the introduction it was emphasised to the children that there were no right or wrong answers and that it was acceptable for them to say "I don't know" to any of the questions they were asked.

2. Practice interview

On the basis of the research findings cited in Chapter 2 and guidance from Home Office (2002), a practice interview was included to further establish rapport and to inform child participants of each stage of the interview, the level of information required of them and also to allow them to practice using the "I don't know" prompts (Schwartz-Kenney, Bottoms and Goodman, 1996).

Children were asked a free recall question about what their teacher looked like⁶². Children were then introduced to the prompts. They were told that the prompts were “pictures and words that can be used to answer questions with”. The “not sure” verbal and the “?” visual prompts were explained. Children were then asked to choose a verbal prompt followed by a visual prompt for a feature element relating to their teacher. It was emphasised to the children that they had not reported something wrong if they were asked the same question more than once, in order to ensure that children did not assume that any previous response they gave was incorrect and, because of this altered their response (Siegal, 1991).

3. Target Description Interview

Free description of target

Children were then asked a free recall question (“Tell me everything that you can remember about what the person in the video looked like”) about what they could remember about the video they had seen the previous day.

Selection of prompts

The free description was followed by two prompted interview conditions (the order of the prompts was counterbalanced across participants). In the visual prompt condition, participants were asked to “Select the best picture which looks like the man’s <insert the name of the feature element>” and again, each choice of prompt (along with the “?” prompt) was pointed out along with its corresponding number of child participants. In the verbal prompt interview condition, child participants were asked “What is the best word for the man’s <insert the name of the feature element>”. Each choice of prompt (along with the “not sure” prompt) was read out loud by the interviewer along with its

⁶² A child’s teacher was chosen as a face familiar to both the child and experimenter but was not present in the interview room.

corresponding number to child participants⁶³. Children were not provided with any feedback about the accuracy of their selections.

4. Identification task

Once the descriptions (i.e. both the visual and verbal prompting sessions) were complete, children were asked to select the target from an array of nine faces. Children were told that they were going to be shown some photos and they had to tell the interviewer “whether or not there was a picture of the man from the video” and if there was a picture of him could they “point to him”. Children were given fair line-up instructions specifying that the target may or may not be present. Children’s responses were recorded on the response sheet. Lastly, children were asked whether they preferred the visual or verbal prompts or if they liked them both the same, and to state their reasons why.

5. Discussion and Closure

At the end of the interview, children were thanked, their well-being checked and they were asked if they had any questions about what they had just done. Children were then returned to their classroom and were asked not to discuss the interview. With the consent of the children’s parents all of the interviews were audio taped. All interviews were conducted by the author who was unfamiliar to the children. The interviews with the child participants lasted between 20 and 40 minutes.

Finally, all results were displayed in a ‘child friendly’ format and sent to each school at the end of the study.

⁶³ Some child participants (particularly older children) asked the interviewer if they could read the verbal prompts themselves. The interviewer read the verbal prompts for at least the first two facial features (Eyebrows and Eyes) to every child and then allowed any child to read the prompts themselves if they wished to. Child participants who read the prompts themselves were encouraged to ask the interviewer for help with any words they were not able to read.

7.3.2 RESULTS

This section considers the child participants' responses and is organised by the aims of the study. It begins with a description of the treatment of the raw data. Additional results from adult participants from Experiment 3 will be compared to child participants' data.

Treatment of Raw Data

Children's selection of the prompts was analysed in terms of accuracy and inaccuracy using the methods described in Section 7.2.2. Children's free descriptions contained very few facial details and therefore could not be analysed in as much detail as was initially intended. For example, the accuracy and inaccuracy of children's selection of the prompts could not be compared to their performance on the free recall task as has been conducted in previous research (e.g. Schwartz-Kenney, Bottoms and Goodman, 1996).

Effects of Target and Prompt Order

Various mixed ANOVA designs were employed to explore the effects of Target and the presentation of Prompt Order on the quantity, accuracy and inaccuracy of children's descriptions were analysed. Results showed some statistically significant main effects and interactions. Effects of Target and Prompt Order are not considered further in this results section. Full ANOVA calculations and explanations are provided in Appendix IVe and Target is also briefly discussed in Section 7.4.

The Effect of Prompts on the Quantity of Children's Descriptions of Unfamiliar Faces (aim (i))

This section considers three levels of description (Free, Visual Prompt and Verbal Prompt). The number of descriptions provided by children, regardless of accuracy, using the prompts were calculated (with the selection of the visual prompt '?' and the

verbal prompt “not sure” removed). Table 7.5 shows a summary of the quantity of descriptions provided by children in each section of the interview.

Age	Section of Interview	Number of descriptions		
		Min	Mean (S.D.)	Max
6 years	Free	0	4.53 (1.93)	11
	Visual Prompt	16	20.97 (2.06)	23
	Verbal Prompt	12	19.57 (2.50)	23
8 years	Free	1	5.70 (2.48)	14
	Visual Prompt	18	22.03 (1.33)	23
	Verbal Prompt	14	21.23 (2.33)	23
10 years	Free	4	7.23 (1.96)	12
	Visual Prompt	17	22.13 (1.53)	23
	Verbal Prompt	12	21.70 (2.34)	23

Table 7.5: Summary of the quantity of descriptions provided by children in Experiment 4

Table 7.5 shows that few free descriptions were provided. The mean number of prompted descriptions provided (visual and verbal) increased with age. The minimum, mean and maximum numbers of descriptions provided by all age groups were higher in both the visual and the verbal prompted sections of the interview than in the free description section, and this was true for all individual child participants. Additionally, the mean number of visually prompted descriptions was higher than the mean number of verbally prompted descriptions for all age groups.

It was hypothesised that using prompts would lead to a statistically significant increase in the quantity of children descriptions. A 3x3 mixed ANOVA design was employed. The between participants factor Age had 3 levels (6-, 8- and 10-years). The within participants factor Description had three levels (Free, Visual Prompt and Verbal Prompt). Results showed the 2-way interaction between Age and Description was statistically non-significant [$F(4,174)=1.819$, $p=0.127$]. There was a statistically significant main effect of Age [$F(2,87)=15.115$, $p<0.0005$, partial $\eta^2=0.258$] suggesting that older participants provided more descriptions. There was also a statistically

significant main effect of Description [$F_{(1.75,152.51)}=2074.823$, $p<0.0005$, partial $\eta^2=0.960$]⁶⁴, supporting the hypothesis stated above.

The statistically significant effect of Description was analysed further using Helmert planned comparisons. The results revealed a statistically significant difference between the mean effect of free descriptions versus the mean effect of the prompts (visual and verbal) [$F_{(1,87)}=3097.130$, $p<0.0005$]. In addition, the comparison of the mean effect of the visual prompt versus the mean effect of the verbal prompt was statistically significant [$F_{(1,87)}=15.106$, $p<0.0005$]. These statistically significant results, together with the data described above, suggest that participants provided statistically significantly more descriptions in the prompted description sections (visual and verbal) of the interview than in the free description section of the interview and statistically significantly more descriptions using the visual prompts compared to the verbal prompts.

The statistically significant main effect of Age was analysed further using a post-hoc Scheffé test, the results of which are shown in Table 7.6.

Age 1	Age 2	Mean Difference	Significance
6 years	8 years	-1.300	$p<0.005$
	10 years	-2.000	$p<0.005$
8 years	6 years	1.300	$p<0.005$
	10 years	-0.700	$p=0.172$
10 years	6 years	2.000	$p<0.005$
	8 years	0.700	$p=0.172$

Table 7.6: Post-hoc Scheffé test of effect of Age on the quantity of descriptions in Experiment 4

⁶⁴ Mauchly's test of sphericity was statistically significant [$W=0.859$, $df=2$, $p<0.005$], so the Greenhouse-Geisser epsilon correction was applied to the degrees of freedom and subsequent p-value.

From Table 7.6 it can be seen that the quantity of descriptions provided by 6-year-olds were statistically significantly different from the quantity of descriptions provided by 8- and 10-year-olds. These results suggest that 8- and 10-year-old children provided statistically significantly more descriptions than the 6-year-old children.

The Effect of Prompts on the Content of Children's Descriptions of Unfamiliar Faces (aim (ii))

The Effect of Prompts on the Content of Children's Free Recall

The descriptions provided by the children in the free recall section of the interview, regardless of accuracy, were categorised into three topic areas (based on categories employed by Greenstock and Pipe, 1997): (1) Person - the physical appearance of the person who conducted the event; (2) Action - the procedures involved during the event; (3) Context - the appearance of the room where the event took place.

Table 7.7 shows the mean number of free descriptions provided by children in each of these categories.

Age	Free Descriptions Mean (S.D)		
	Person	Action	Context
6 years	2.77 (1.72)	1.7 (0.7)	0.07 (0.25)
8 years	3.6 (2.06)	1.93 (0.64)	0.17 (0.46)
10 years	5.33 (1.83)	1.83 (0.53)	0.07 (0.25)
All	3.9 (2.14)	1.82 (0.63)	0.1 (0.34)

Table 7.7: Mean number of free recall descriptions provided by children in Experiment 4

As expected (due to the preparation of the videos as facial stimuli) on average, the highest number of total descriptions were provided for person related details, followed by action descriptions and the lowest number of descriptions provided were provided for context information. This was true for all age groups of children.

Additional analyses of *person* descriptions further divided descriptions by (i) face/body (ii) clothes and (iii) other (e.g. subjective details such as the target being “grumpy”, “angry”) Table 7.8 shows this information.

Age	Person Topics			
	Mean (<i>S.D</i>)			
	Face	Body	Clothes	Other
6 years	1.63 (1.73)	0.03 (0.18)	0.27 (0.52)	0.83 (0.79)
8 years	2.93 (1.91)	0.07 (0.37)	0.03 (0.18)	0.57 (0.9)
10 years	4.8 (1.75)	0 (0)	0.17 (0.38)	0.37 (0.76)
All	3.12 (2.21)	0.03 (0.23)	0.16 (0.39)	0.59 (0.83)

Table 7.8: Mean number of free recall person descriptions provided by children in Experiment 4

Table 7.8 indicates that the mean number of person descriptions increased with age. On average, the highest number of person descriptions were for facial details, followed by other person descriptions, clothing information and body information, for all age groups of children. The mean number of facial descriptions increased with age, whereas this pattern was reversed for clothing and for other descriptions.

Further analyses of children’s free descriptions were not possible due to the limited number of free descriptions of the targets by all children. Therefore, the following results sections do not include the free description level of the within participants factor Description. Instead, analyses will include the within participants factor Prompt with two levels (Visual and Verbal).

The Effect of Prompts on the Accuracy Content of Children’s Prompted Recall

Although the use of prompts statistically significantly increased the quantity of information provided, the accuracy of the increased information has not yet been considered.

The accuracy of each child participant's selection of prompts was calculated as a percentage (as described in Section 7.2.2)⁶⁵. Table 7.9 shows the mean accuracy scores for child participants' selection of prompts.

	Accuracy (%)	
	Mean (S.D)	
Age	Visual Prompt	Verbal Prompt
6 years	46.23 (11.80)	44.14 (15.30)
8 years	55.85 (11.97)	53.35 (14.44)
10 years	62.82 (12.08)	58.58 (14.44)
All	54.97 (13.65)	52.03 (15.75)

Table 7.9: Mean accuracy scores for children in Experiment 4

Table 7.9 shows that the mean accuracy scores for both visual and verbal prompt conditions increased with age. The mean accuracy scores for the visual prompt condition were higher than the corresponding mean scores for the verbal prompt condition for all age groups and the largest difference between these two scores was for 10-year-olds. Additionally, the standard deviations were lower for the visual prompt condition scores than for the verbal prompt condition scores, for all ages.

A 2x3 mixed factorial ANOVA design was employed. The within participants factor Prompt had two levels (Visual and Verbal), and the between participants factor Age had 3 levels (6-, 8- and 10-years). The dependent variable was the accuracy score. The 2-way interaction between Prompt and Age was statistically non-significant [$F_{(2,87)}=0.317$, $p=0.729$]. There was a statistically significant main effect of Prompt [$F_{(1,87)}=6.354$, $p<0.05$, partial $\eta^2=0.068$], suggesting that descriptions elicited using the visual prompts were awarded higher accuracy scores. There was also a statistically significant main effect of Age [$F_{(2,87)}=12.272$, $p<0.0005$, partial $\eta^2=0.220$], suggesting that descriptions provided by older participants were awarded higher accuracy scores.

⁶⁵ As described in Section 7.2.2, the visual comparison of target photographs with the visual prompts by independent judges was explored as an alternative measure of accuracy. This measure showed similar patterns of results and therefore the findings are not presented here.

The statistically significant main effect of Age was further analysed using a post-hoc Scheffé test, the results of which are shown in Table 7.10.

Age 1	Age 2	Mean Difference	Significance
6 years	8 years	-9.416	$p < 0.05$
	10 years	-15.512	$p < 0.005$
8 years	6 years	9.416	$p < 0.05$
	10 years	-6.096	$p = 0.161$
10 years	6 years	15.512	$p < 0.005$
	8 years	6.096	$p = 0.161$

Table 7.10: Post-hoc Scheffé test of the effect of Age on the accuracy of descriptions in Experiment 4

From Table 7.10 it can be seen that mean accuracy scores for descriptions provided by 6-year-olds were statistically significantly different from mean accuracy scores for descriptions provided by 8- and 10-year-olds. These results, together with the data described suggest that 8- and 10-year-old children provided statistically significantly more accurate descriptions than the 6-year-old children.

Effect size was calculated and showed that only 6.8% of the overall variation in accuracy was attributable to the influence of Prompt.

The Effect of Prompts on the Inaccuracy Content of Children's Prompted Recall

The inaccuracy of each child participants' selection of prompts was calculated as a percentage (as described in section 7.2.2). Table 7.11 shows the mean inaccuracy scores for child participants' selection of visual and verbal prompts.

Age	Inaccuracy Mean (S.D)	
	Visual Prompt	Verbal Prompt
6 years	24.87 (11.64)	27.20 (14.93)
8 years	18.73 (11.24)	21.90 (12.81)
10 years	13.40 (9.38)	17.17 (12.84)
Total	19.00 (11.67)	22.09 (14.03)

Table 7.11: Mean inaccuracy scores for children in Experiment 4

Table 7.11 shows a similar pattern of results to the mean accuracy scores in Table 7.9. The mean inaccuracy scores for both the visual and verbal prompt conditions decreased with age. The mean inaccuracy scores for the visual prompt condition were lower than the corresponding mean inaccuracy scores for the verbal prompt condition and the largest difference was for 10-year-olds. Additionally, the standard deviations were lower for the visual prompt condition scores than for verbal prompt condition scores, for all ages.

A 2x3 mixed ANOVA was employed on the same factors as the previous analyses with the inaccuracy score as the dependent variable. The 2-way interaction between Prompt and Age was statistically non-significant [$F_{(2,87)}=0.103$, $p=0.903$]. There was a statistically significant main effect of Prompt [$F_{(1,87)} = 5.669$; $p<0.05$, partial $\eta^2=0.061$], suggesting that descriptions elicited using the visual prompts were awarded lower inaccuracy scores. There was also a statistically significant main effect of Age [$F_{(2,87)}=7.719$, $p<0.005$, partial $\eta^2=0.152$], suggesting that descriptions provided by older participants were awarded lower inaccuracy scores. The statistically significant main effect of Age was further analysed using a post-hoc Scheffé test, the results of which are shown in Table 7.12.

Age 1	Age 2	Mean Difference	Significance
6 years	8 years	-5.72	$p=0.119$
	10 years	-10.75	$p<0.005$
8 years	6 years	5.72	$p=0.119$
	10 years	-5.03	$p=0.191$
10 years	6 years	10.75	$p<0.005$
	8 years	5.03	$p=0.191$

Table 7.12: Post-hoc Scheffé test of the effect of Age on the inaccuracy of descriptions in Experiment 4

From Table 7.12 it can be seen that mean inaccuracy scores for descriptions provided by 10-year-olds were statistically significantly different from mean inaccuracy scores for descriptions provided by 6-year-olds. This result, together with the data described suggest that 6-year-old children provided statistically significantly more inaccurate descriptions than 10-year-old children.

Again, effect size was calculated and showed that only 6.1% of the overall variation in inaccuracy was attributable to the influence of Prompt.

The Effect of Prompts on the Time Taken to Obtain Prompted Descriptions from Children (aim (iii))

The mean time taken by children to select the visual and the verbal prompts was calculated by listening to and timing the tape recorded interviews. Table 7.13 shows this information.

Age	Prompts		Mean (S.D)
	Verbal (S.D)	Visual (S.D)	
6 years	8:37 (1:33)	5:51 (1:06)	7:14 (1:56)
8 years	7:12 (1:26)	5:32 (1:01)	6:12 (1:29)
10 years	5:30 (1:19)	4:46 (1:22)	5:08 (1:23)
All	7:07 (1:55)	5:23 (1:15)	6:15 (1:50)

Table 7.13: Mean time taken (minutes:seconds) by children to select prompts in Experiment 4

Table 7.13 shows that the mean time taken to select the visual and verbal prompts decreased with age. The mean time taken to select the visual prompts was lower than the mean time taken to select the verbal prompts for all ages and the largest difference between these two scores was for 6-year-olds. Finally, the range of mean times taken for to select the visual prompts was smaller than the range of mean times taken to select the verbal prompts.

A 2x3 mixed ANOVA design was employed on the same factors as the previous analyses. There was a statistically significant main effect of Prompt [$F_{(1,87)} = 164.202$;

$p<0.0005$, partial $\eta^2=0.654$], suggesting that descriptions were elicited faster using the visual prompts. There was also a statistically significant main effect of Age [$F_{(2,87)}=25.312$, $p<0.0005$, partial $\eta^2=0.368$] suggesting that older participants were faster selecting the prompts. However, there was also a statistically significant 2-way interaction between Prompt and Age [$F_{(2,87)}=19.058$, $p<0.0005$] on the time taken. Figure 7.7 illustrates the interaction.

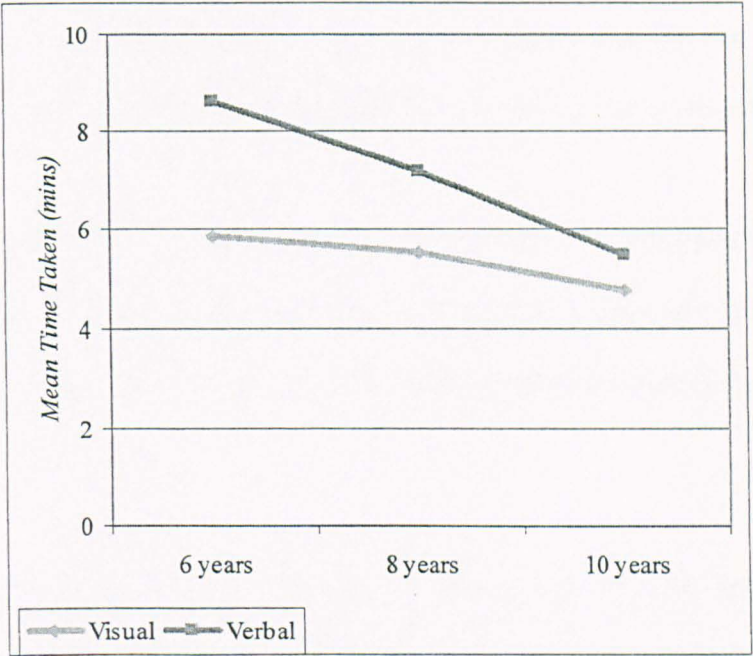


Figure 7.7: Mean time taken to select prompts by children in Experiment 4

Figure 7.7 shows that the interaction was due to the relatively small difference between age groups for the visual prompts but a large difference between age groups for the time taken to select the verbal prompts. *All* of the 6 year olds took less time to select the visual prompts than the verbal prompts however, four of the 8 year olds and eight of the 10 year olds spent more time selecting the visual prompts than the verbal prompts.

Children’s Use of the “don’t know” Prompts (aim (iv))

Table 7.14 summarises children’s selection of the visual ‘?’ and verbal “not sure” prompts.

Age	Visual ‘?’			Verbal “not sure”		
	min	mean (S.D)	max	min	Mean (S.D)	max
6 years	0	2.03 (2.06)	7	0	3.43 (2.50)	11
8 years	0	0.97 (1.33)	5	0	1.77 (2.33)	9
10 years	0	0.87 (1.53)	6	0	1.30 (2.34)	11
All	0	1.29 (1.77)	7	0	2.17 (2.54)	11

Table 7.14: Children’s selection of the “don’t know” prompts in Experiment 4

Table 7.14 shows that the use of visual ‘?’ and verbal “not sure” decreased with age (with the 6-year-olds using these approximately twice as much as the 8- and 10-year-olds). The use of the verbal “not sure” prompt was higher than the use of the visual ‘?’ prompt for all age groups and the largest difference was for the 6-year-olds.

Adult participants also used these prompts, even though the target photograph remained present. Like the child participants, adults used the verbal “not sure” prompt more than the visual ‘?’ prompt (an average of 0.75 and 0.68 times for verbal and visual prompts respectively).

For children the highest use of the visual ‘?’ prompt was for the feature element Ear Shape (means=0.23, 0.16 and 0.16 for 6, 8 and 10 year olds respectively) and this was also true for adults (mean=0.32). This may have been due to the fact that part of Target 2’s ear was obscured by hair. The highest use of the verbal “not sure” prompt was for a number of features: Ear Setting and Eye Shape for 6 year olds (mean= 0.3), Eye Colour, Mouth Shape, Mouth Size, Hair Length and Nose and mouth spacing (0.13) for 8 year olds and Eyebrow Shape, Eye Colour, Nose Tip, and Nose, mouth and chin spacing (0.1) for 10 year olds. For adults the highest use of verbal “not sure” prompt was again for Ear Shape (0.29) possibly for the reason previously provided.

Children's Performance at the Identification Task Involving the Target Described (aim (v))

Once the descriptions (i.e. both the visual and verbal prompting sessions) were obtained children were asked if they could select the target from an array of 9 faces. Figure 7.8 shows the percentage of children who made correct and incorrect identifications (the probability of a correct identification is 11% and is illustrated by a line).

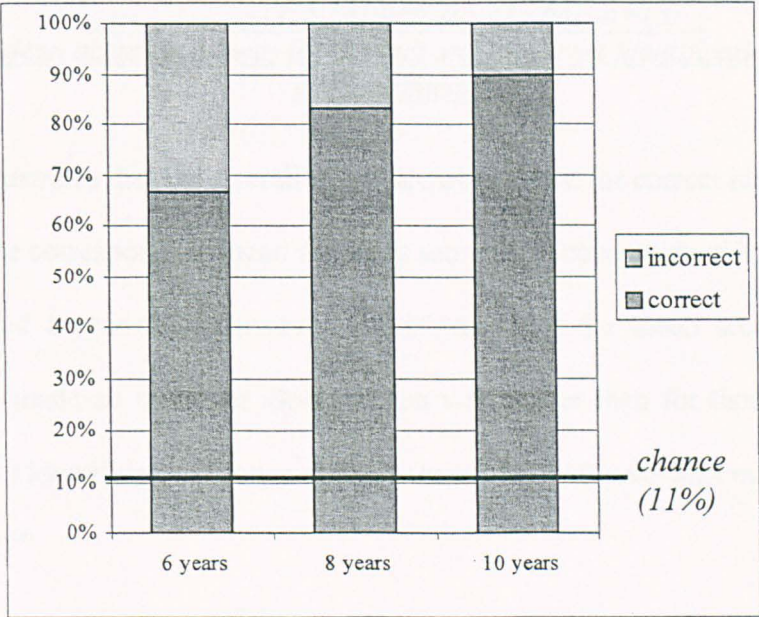


Figure 7.8: Percentage of correct and incorrect identifications by children in Experiment 4

Figure 7.8 shows that the percentage of correct identifications increased with age. Even though 6 year olds provided correct identifications approximately 25% less than 10 year olds they were still over 50% above chance. Although there was a difference between age, a Chi-Square test showed that the relationship between overall identification result and age of child participant was statistically non-significant [$X^2=0.067$, $df=2$, $p=0.069$].

Using identification performance as the independent variable, the mean accuracy scores (based on the accuracy calculations using adult participants' responses described in Section 7.2.2) for both correct and incorrect identifications was calculated, this information is shown in Table 7.15.

Age	Mean Accuracy Score (%) (S.D) (n)	
	Correct Identification	Incorrect Identification
6 years	47.06 (12.36) (n=20)	41.45 (11.99) (n=10)
8 years	56.58 (12.07) (n=25)	44.74 (10.13) (n=5)
10 years	60.39 (12) (n=27)	63.51 (12.81) (n=3)
<i>All</i>	55.36 (13.13)(n=72)	46.04 (13.64) (n=18)

Table 7.15: Mean accuracy scores for correct and incorrect identifications by children in Experiment 4

Table 7.15 illustrates that the overall mean accuracy score for correct identification was higher than the corresponding mean accuracy score for incorrect identification. This was true for 6- and 8-year-olds. However, for 10-year-olds the mean accuracy score for children who made an incorrect identification was higher than for those children who made a correct identification (however, only three of the 10-year-olds made an incorrect identification)⁶⁶.

Additional Observations

Preferences

At the end of the interview child participants were asked whether they found the visual or verbal prompts (or both) the same, easier and/or more helpful when they had to describe features of the man's face. Figure 7.9 depicts 6-, 8-, and 10-year olds' preferences.

⁶⁶ Due to the unequal numbers for correct and incorrect identifications no inferential statistics were conducted.

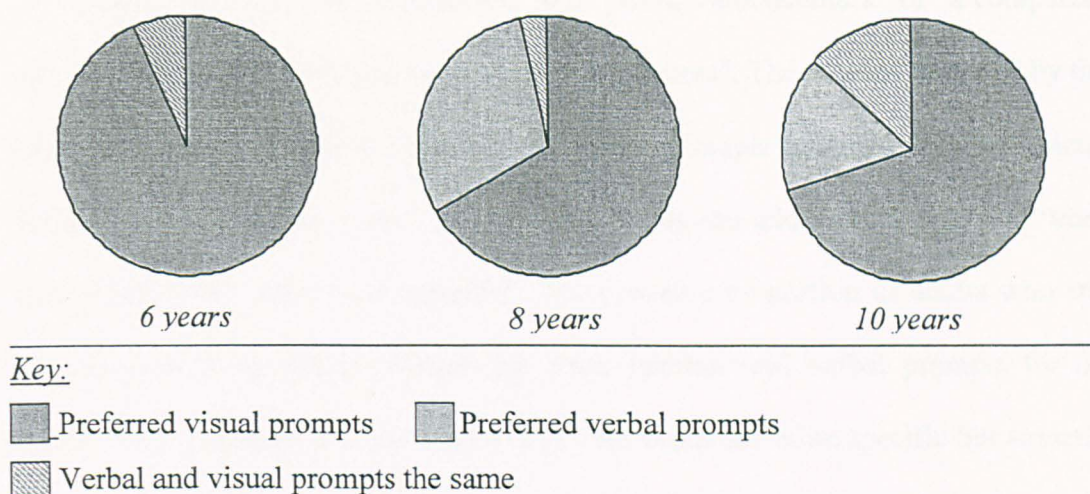


Figure 7.9: 6-, 8-, and 10- year olds preferences for prompts in Experiment 4

Figure 7.9 shows the majority of children in each age group had a preference for visual prompts when describing the target face. This was highest for the 6 year olds where 93% of children had a preference for visual prompts; 66% of 8 year olds and 70% of 10 year olds preferred visual prompts. None of the 6-year-olds preferred the verbal prompts.

One 10-year-old stated that the visual prompts were better because “you could see what you were pointing to and saying, and with the words you couldn’t see what you mean”. Another 10 year old stated that they “didn’t know words...I don’t know what oval means”. Other quotes include 10 year olds stating that they preferred the visual prompts because “with the words, didn’t know how wavy or how straight things were” and “the words couldn’t...you had to picture them in your mind”.

Additionally, a large proportion of adults (60%) preferred the visual prompts (30% preferred the verbal prompts). The reasons provided for adults preferring the visual prompts include that “[descriptions] are very difficult to put into words”, “people have a very individual sense of what they mean” and “with pictures you can make some sort of

comparison” they are “self explanatory” and “provide a benchmark” or “a comparison” and that there is “less personal bias involved in pictures”. The reasons provided by those adults who stated that they preferred the verbal prompts included that the “pictures limited...words describe more” and that the “words can add more detail” and “can be more descriptive” and “more specific”. There were a proportion of adults who stated that they preferred visual prompts for some features and verbal prompts for other features. For example, one adult stated that “the words are more specific but sometimes picture make it clearer, for example, face shape was easier in pictures as it is difficult to describe in words” and that the “spacing easier in pictures” (referring to the configural prompts). Another adult stated that “can use pictures and elaborate with words”.

7.4 DISCUSSION OF STUDY 3

Analyses of the results initially looked at the effect of prompts on the quantity of descriptions provided by children, regardless of the accuracy of that information. Analyses then went on to consider the accuracy of the prompted information provided. Finally, the time taken to obtain the prompted descriptions, children’s use of “don’t know” prompts and children’s performance at the identification task involving the target described was considered.

7.4.1 DISCUSSION OF STUDY 3 RESULTS

The Effect of Prompts on the Quantity of Children’s Descriptions of Unfamiliar Faces (aim (i))

When looking at the quantity of children’s descriptions, the results from Study 3 indicate that when asked to provide a free recall of a target’s appearance older children provided statistically significantly more descriptions than younger children. However, children of all ages were unable to freely recall information, especially information

useful for composite construction. When provided with prompts the hypothesis was supported with children of all ages providing statistically significantly more information. Even children who provided no free recall descriptions provided information with the prompts. These findings also replicate the findings from Study 2, indicating again that children require specific prompting when being asked to provide facial descriptions. The fact that children could not freely retrieve information about the target's appearance suggests that communicating the information might have been difficult given their limited language development (as described in section 2.1.3). For example, one 8-year-old stated in the free recall section that "I know lots, I can remember lots but I can't describe".

On average, children provided a statistically significantly higher number of descriptions when using the visual prompts than the verbal prompts (regardless of accuracy) thus suggesting that the linguistic skills needed to use the visual prompts were not as sophisticated or complex as those needed for accurate verbal recall (Gabarino and Stott, 1992; Pipe *et al.*, 2002).

Furthermore, the data showed that the quantity of descriptions provided increased with age which is unsurprising given that it is usual to find age difference in the completeness of descriptions (see review by Ceci and Bruck, 1993). In the present study, 8- and 10-year-olds provided statistically significantly more descriptions than 6 year olds. This finding is consistent with the heuristic identified in the literature that age 7-years represents a significant shift in children's abilities and associated developmental changes (Westcott *et al.*, 2002).

The Effect of Prompts on the Content of Children's Descriptions of Unfamiliar Faces (aim (ii))

The majority of children's free descriptions were made up of person descriptions, suggesting that children had attended to the target (this was further supported by the high percentage of correct identification rates). The total number of person descriptions increased with age, consistent with research that children's facial processing skills improve within the first decade of life (for review see Chung and Thompson, 1995). Although the highest numbers of descriptions were provided for the face, the number which would be useful when constructing a composite with children was low.

Accuracy for the visual prompt condition was higher than for the verbal prompt condition for children of all ages. Additionally, inaccuracy for the visual prompt condition was statistically significantly lower than inaccuracy for the verbal prompt condition for children of all ages as would be expected on the basis of the language research findings cited in Section 2.1.3.

The accuracy of prompt selection increased with age and the inaccuracy of children's selection of the prompts decreased with age. Such developmental differences are not surprising given research demonstrating changes in memory development cited in Section 2.1.2. 10-year-olds provided the most accurate, and least inaccurate, prompted descriptions and 6-year-olds provided the least accurate and most inaccurate descriptions. This is consistent with research conducted by Schwartz-Kenney, Bottoms and Goodman (1996) who reported that their visual line-up techniques seemed to hold most benefits for older children (8- to 9-years) than for younger children (5- to 6-year-olds).

In some cases where children were inaccurate in their prompted recall of features, children may have chosen the 'normal' exemplar (for example 'average' length nose, or 'average' size eyes) rather than the 'correct' exemplar, as they have been shown to describe how things 'normally happen' (Jones, 2003). This is supported by the findings from Study 2. For example, in Study 2, for some feature elements children tended to not use varying descriptors (Pattern 2). Again, for the feature element Face shape, children tended to use the descriptors oval and round. Future research could apply the patterns from Study 2 to children's selection of the prompts. With some prompts children may not have wanted to select picture or words which are not 'normal'.

Finally, although there was a significant main effect of Prompt the majority of the variation in accuracy and inaccuracy could not be accounted for (only 6.8% of the overall variation in accuracy, and only 6.1% of the overall variation in inaccuracy, was attributable to the influence of Prompt). This highlights the large variations within the data for each age group.

The Effect of Prompts on the Time Taken to Obtain Prompted Descriptions from Children (aim (iii))

Children were statistically significantly faster when selecting the visual prompts than the verbal prompts and older children were statistically significantly faster than younger children when selecting the prompts. There was a small difference between ages for the time taken to select the visual prompts but a large difference between age groups for verbal prompts. The largest difference was for 6-year-olds due to the fact that all 6-year-olds took less time to select visual prompts whereas there were large individual differences within the 8- and 10-year-old groups. Many of the 6-year-olds also required the prompts to be read out to them by the interviewer whereas fewer numbers of the 8- and 10-year olds required this. Additionally, a number of children across the age groups

asked for further explanations of the verbal prompts which increased the time spent selecting them.

The average time spent with witnesses selecting the prompts was under nine minutes. In response to the questionnaire survey (Study 1), operators reported spending around 40 minutes on average obtaining children's facial descriptions (exclusive of rapport building) and throughout the survey operators raised concerns about the time they spent interviewing children and related to this, problems with children's concentration levels. The results from the present study indicate that the visual prompts (and indeed the verbal prompts) reduced the time spent interviewing children compared to operators' responses in Study 1, and that in this shorter interview time, accurate facial descriptions could be obtained from children which were of potential use for subsequent composite construction.

A further observation made by the interviewer was that at the end of the task some children (particularly younger children) were losing concentration. However, in a forensic setting children would only be asked to use one set of prompts and therefore the total time spent would not be as long as in the present study.

Children's Use of "don't know" Prompts (aim (iv))

The use of the visual and verbal don't know options decreased with age, possibly as children's recall for the target increased with age (Chung and Thompson, 1995). All age groups used the verbal "not sure" prompt more than the visual '?' prompt. This finding suggests that children may have understood the visual prompts more than the verbal prompts, thus providing further support for the age appropriateness of the visual prompts suggested above.

Concerns about children answering option posing questions were reviewed in Section 2.2.2 (e.g. Poole and White, 1993). For example, children interpreting option-posing questions as calling for a response, even when they have no idea what the question is asking. These concerns raised the issue of the need for some sort of “don’t know” option, which was confirmed by findings reported by Schwartz-Kenney, Bottoms and Goodman (1996) and Clark *et al.* (2001). Therefore, of particular importance was the finding that children appeared to be comfortable selecting both visual and verbal “don’t know” prompts rather than attempting to answer questions which they did not fully understand (Saywitz and Synder, 1993) or could not remember the answer to. The findings also provide support for the inclusion of a practice task of the use of these prompts (Schwartz-Kenney, Bottoms and Goodman, 1996).

Children’s Performance at the Identification Task Involving the Target Described (aim (v))

In the identification task all children performed at above chance and there was no statistically significant difference between different ages of children. These findings are consistent with a meta analysis conducted by Pozzulo and Lindsay (1998), who found that children aged 5-years and over did not differ significantly from adults with regard to correct identification rate. The percentage of correct identifications for all age groups were higher than predicted by Chance and Goldstein (1984) who showed that the number of correct identifications appear to increase with age from 50 to 60% for 6- to 8-year-olds to 60 to 70% for 9- to 11-year-olds up to 70 to 80% for adults. Such high levels of identification could possibly be due to the effect of providing both verbal and visual description. Studies have demonstrated the positive effects of verbal description, rehearsal and elaboration on the later recognition of faces (e.g. Yu and Geiselman, 1993).

The accuracy scores (based on the accuracy calculations using adult participants' responses) for correct identifications were statistically significantly higher than for incorrect identifications for 6- and 8-year-olds. However, the opposite is true for 10-year-olds which was probably due to the low number of 10-year-old children making an incorrect identification.

The key point is that children who did not make a correct identification were still scoring fairly high with the accuracy of the selection of prompts. A possible explanation is that verbally describing the target may have overshadowed children's abilities to correctly identify the target (Schooler and Engstler-Schooler, 1990): working through participants' descriptions feature by feature may have destroyed the configuration of the target face in the participant's memory, with the result that a participant may have a fairly accurate description but not being able to correctly identify the target, particularly for younger children. This is consistent with research cited in Section 3.2.2, that configural processing becomes more resilient with age, especially after 6-years of age (e.g. Donnelly and Hadwin, 2003). Perhaps a live identification coupled with a recovery period would improve children's correct identification performance.

A note-worthy procedural point was that during the identification task, the interviewer had to explain to children (especially younger children) that if the target was present in the line-up that it would be a "picture of the man at another time". This is consistent with research that younger children are poor at deciding whether two simultaneously presented pictures are of the same person if the photographs differ in terms of their lighting, angle of viewpoint, facial expression or the clothing worn (Diamond and Carey, 1977; Flin, 1980).

Summary of Study 3 Results

Overall, children were statistically significantly more accurate, less inaccurate and faster when selecting the visual prompts compared to the verbal prompts to describe an unfamiliar target. Although there were no large differences between accuracy and inaccuracy in the visual and verbal prompt conditions, importantly, the visual prompts never resulted in lower accuracy or higher inaccuracy than the verbal prompts. Furthermore, this positive effect of visual over verbal prompts was based on verbal prompts constructed from children's descriptions, which were therefore "child friendly" or "age appropriate" verbal prompts. Finally, the majority of children preferred the visual prompts.

In summary, all of the results showed a consistent and significant improvement in the quantity of descriptions (from free to prompted) and the accuracy and inaccuracy of prompted descriptions with age, suggesting a quantitative rather than a qualitative shift with age.

7.4.2 LIMITATIONS OF STUDY 3

Due to restrictions from the school used in the present study, audio recording was used, supplemented by notes made by the author throughout each interview. A more accurate measure could be taken by video recording interviews. ABE (Home Office, 2002) also stresses that interviews with child witnesses should be video recorded and following this guidance could make the interview process forensically more valid. This is of particular importance when using prompts with children. Children may not verbalise their choice of prompt and a video would illustrate the child's selection and could not be open to scrutiny if presented to a court as evidence.

Another limitation of Study 3 surrounds the measures of accuracy and inaccuracy. Even with the targets remaining in view there was a large amount of variability in adults' selection of the prompts therefore the initial measure of accuracy which was proposed (based on adults' modal selection of the prompts) could not be used. Therefore, alternative measures of accuracy had to be explored.

A further limitation of the present study was that the accuracy of children's free recall was not measured. Therefore, the accuracy (and inaccuracy) of children's selection of the prompts could not be compared to their performance on the free recall task as has been conducted by previous research (e.g. Schwartz-Kenney, Bottoms and Goodman, 1996). However, this was due to the very limited number of facial descriptions provided by children in the free recall interview section.

Finally, as with any research of this kind, due to the artificial nature of many aspects of the experimental environment employed the methodology of this study may be questioned on the grounds of its ecological validity. This will be considered in more detail in the main discussion of the thesis (Chapter 9).

7.4.3 SUMMARY: IMPLICATIONS OF STUDY 3 AND FUTURE RESEARCH

A number of issues have been raised which have implications for future research. As described, the main aim of the identification task was to assess whether children had attended to the target. Further research could determine the impact of target-absent line-ups. In particular, any effect of verbal overshadowing (e.g. Schooler and Engstler-Schooler, 1990) or image (visual) overshadowing (Clark, 2000) of using the prompts on subsequent identification could be investigated. This issue is beyond the scope of the current thesis.

The use of two target faces resulted in some statistically significant effects on the accuracy and inaccuracy of the data. Although the use of more than one target face is a positive design consideration, the next study attempted to overcome this limitation by using more than two target faces.

Finally, the use of the prompts (particularly the visual prompts) may potentially enhance the communication process between a child witness and an operator when constructing a composite. Prompts may assist operators with the issues of time and language raised by operators in response to the questionnaire survey (Chapter 5) and with the issues described at both the level of the operator and the witness described in Section 4.2.2. Therefore, the main issue to be considered in terms of the current thesis is the careful development of the use of prompts with children when constructing a composite. It is this issue which was used to direct the following studies in the current thesis.

CHAPTER 8: STUDY 4

THE EFFECT OF PROMPTS ON CHILDREN'S FACIAL COMPOSITES OF UNFAMILIAR FACES

This chapter describes Study 4, an empirical investigation of sets of visual prompts and verbal prompts as an appropriate interview technique to assist adult and child participants' composite constructions of unfamiliar faces. Study 4 comprised three experiments: Experiment 5 Composite Construction – this involved child and adult participants' selection of prompts and construction of composites using the computerised system E-FIT; Experiments 6 and 7 Composite Evaluation - Experiment 6 consisted of the subjective evaluation of the composites (using ranking and rating tasks); and, Experiment 7, the objective evaluation of the composites (using a forced choice matching task). A discussion section is included at the end of each experiment.

8.1 INTRODUCTION

Chapter 4 considered the practical aspects of facial composite production, including a comparison of the effectiveness of non-computerised and computerised composite systems. The results cited in Section 4.1.3 showed no decline in the quality of non-computerised composites constructed under target absent conditions versus those constructed in the presence of the target (Davies *et al.*, 1978; Ellis *et al.*, 1978; Laughery and Fowler, 1980). These findings suggested serious limitations in the flexibility of non-computerised systems. However, later research studies conducted with the more flexible computerised system Mac-a-Mug also produced significant target present/target absent effects which could not be attributed to the system (Cutler *et al.*, 1988; Koehn and Fisher, 1997). In their conclusion, Koehn and Fisher (1997) suggested that witnesses' poor performance under memory conditions could be due to the feature

driven composite systems such as Photo-FIT and Mac-a-Mug requiring the selection of individual features in isolation, which is incompatible with adults facial processing (e.g. Young *et al.*, 1987). In addition, contemporary research cited in Section 3.2.2 reported that by the age of 6-years children are also perceiving and remembering faces holistically (e.g. Gilchrist and McKone, 2003). Koehn and Fisher proposed that a composite system which encouraged a more holistic approach to facial construction might yield higher quality images for example, the E-FIT system.

Evaluations of the E-FIT system with adult participants has produced mixed results. As described in Section 4.1.3, Turner (2004) provided some supporting evidence for the positive effects of a holistic approach over a featural approach *within* the E-FIT system under target absent conditions. In contrast, a direct comparison of the computerised E-FIT system and the non-computerised Photo-FIT system by Davies *et al.* (2000) only found an advantage of the E-FIT system under target present conditions. The results of Davies *et al.* imply that the quality of composites are not only due to the composite system itself but are also a function of the task required of the witness. Consistent with this view, Laughery and Fowler (1980) and Koehn and Fisher (1997) suggested that in addition to the need for a more holistic approach to composite construction, the composite interview process should also be considered.

The lack of research studies addressing children's abilities to construct facial composites was documented in Section 4.3.2. Detailed procedural information, the principal results and a brief evaluation of each of the three studies which have been conducted (Davies *et al.*, 1989; Flin *et al.*, 1989; Schwartz-Kenney, Norton *et al.*, 1996) is presented in Appendix Va. Collectively, the results from these studies demonstrated that from the age of 5- to 6-years, children were capable of constructing a composite

using these systems. In each of the three studies conducted the overall likenesses of composites were low for all the children, regardless of age. Flin *et al.* (1989) found a statistically significant difference in the percentage of correctly sorted composites constructed by adults and by children, and the difference between 11- and 8-year-olds approached statistical significance. However, Flin *et al.* stressed that the Photo-FITs produced by the youngest group were still sorted at an overall standard which was markedly above chance and that among these composites were a number of excellent likenesses, thus suggesting that adults do not consistently produce facial composites of a better quality than children. Whilst Davies *et al.* (1989) reported no statistically significant difference between likeness ratings for composites constructed by 6- to 7- and 10- to 11-year-olds, in contrast, Schwartz-Kenney, Norton *et al.* found a statistically significant difference between likeness ratings for composites constructed by 5- to 6- and 8- to 9-year olds. These differences in findings may be explained by the inclusion of the younger age in the latter study. It is important to note that the two studies did not include adult comparison groups and it is necessary to compare children's performance to that of adults (Davies, 1996).

A number of methodological differences may also have contributed to the differences in the findings of the three studies, for example, in each study children were constructing composites under different circumstances. Flin *et al.* (1989) conducted a laboratory study in which participants were shown a photographic stimulus and no delay was included before recall. In contrast, the studies conducted by Davies *et al.* and Schwartz-Kenney, Norton *et al.* were more forensically realistic, using a live interaction. Davies *et al.* also included a one week delay. Additionally, the composites constructed in each study were evaluated under different circumstances. The composites constructed in the studies by Davies *et al.* and Schwartz-Kenney, Norton *et al.* were evaluated using a

subjective measure, requiring judges to rate the likeness of the composites to the target. Flin *et al.* included the objective measure of utility of composites, by asking judges to identify or match the composites to the correct target. Finally, Flin *et al.* showed no significant difference between the photo absent and the photo present condition for all age groups which again appears to reflect the inflexibility of composite systems rather than memory problems of the witness.

In summary, the ability of children to use composite systems is still in question. To date, no research examining children's abilities to construct composites using the computerised system E-FIT have been published. Additionally, research is also needed which focuses not only on the composite system, but on the composite interview process (Koehn and Fisher, 1997; Laughery and Fowler, 1980). As described in Study 3, the visual and verbal prompts were designed in order to improve the interview process between a witness and an operator, by avoiding the problems of language and concentration highlighted by operators' responses in Study 1 and described in Chapters 2 to 4. That is, the prompts were designed to assist witnesses in providing facial descriptions and interpreting operators' prompts and operators in eliciting and interpreting witnesses' descriptions. In addition, the prompts also map directly onto the E-FIT composite system in order to further reduce the reliance on an operator 'translating' witnesses' descriptions to match the E-FIT index.

The findings from Experiment 4 in Study 3 suggested that the use of a set of visual prompts is an appropriate interview technique which can be used to significantly enhance children's facial descriptions, in terms of reducing time taken, and by increasing the number and accuracy of children's descriptions for all age groups of children. The majority of children also stated that they preferred using the visual

prompts. These findings (together with the design considerations of the prompts described in Section 7.1.4) suggest possible benefits of these prompts in composite interviews with children.

To conclude, preliminary research is required to examine children's abilities to construct composites with the computerised system E-FIT, and to investigate the use of the prompts as an appropriate composite interview technique with children. Such research is the focus of Study 4.

8.1.1 RESEARCH AIMS OF STUDY 4

(1) An initial aim of Study 4 was to investigate whether a number of the findings from Experiment 4 would be replicated, in particular, to explore any differences between using a set of visual prompts or a set of verbal prompts on:

(i) the time taken to obtain the prompted descriptions. It was hypothesised that there would be a statistically significant effect of *prompt* (i.e. using the visual prompts would lead to descriptions being elicited statistically significantly faster than using the verbal prompts) and a statistically significant effect of *age* (i.e. that descriptions would be elicited statistically significantly faster from older participants).

(ii) participants' use of the "don't know" prompts. The trends apparent in Experiment 4 were explored: that the use of the "don't know" prompts would decrease with age and the use of the verbal "not sure" prompt would be greater than the use of the visual '?' for all ages of children.

(iii) participants' performance at an identification task following composite construction. It was hypothesised that the effect of Age would be statistically non-significant.

(2) The main aim of Study 4 was to explore the use of the set of visual prompts and verbal prompts on adult and child participants' abilities to construct facial composites with the computerised composite system E-FIT. The study was designed to investigate whether the potential benefits of the prompts observed on children's descriptions in Experiment 4 (Study 3) could be extended to the composite production process, and in particular to determine whether:

(iv) children and adults were able to construct computerised composites using the visual and verbal prompts in an experimental setting. This aim was treated as exploratory and no specific hypotheses were stated.

As part of this main aim, Study 4 was designed to evaluate the composite images produced, in particular to explore any:

*(v) effect of age on the **subjective likeness** of composite images produced.* It was hypothesised that there would be a general improvement in the likeness of composites produced with age.

*(vi) differences between using the set of visual or verbal prompts as a method of obtaining initial description in terms of the **subjective likeness** of composite images produced.*

*(vii) effect of age on the **objective utility** of composite images produced.* It was hypothesised that there would be a general improvement in the utility of composites produced with age.

*(viii) differences between using the set of visual or verbal prompts as a method of obtaining initial description in terms of the **objective utility** of composite images produced.*

8.1.2 COMPOSITE CONSTRUCTION AND EVALUATION CONSIDERATIONS

The composite construction process (Experiment 5) and the subsequent evaluation of the composites (Experiments 6 and 7) utilised an experimental design. This allowed a large scale study to be conducted as an initial first step in what may potentially be a whole series with children. Within this design a time limit of approximately 30-minutes with each participant was set (for the elicitation of a description and composite construction). This time limit was imposed as a result of the schools involved in the study, and because of operators' concerns about children's concentration levels raised in response to the questionnaire survey in Study 1.

During composite construction with participants the seven main facial features were worked through individually in the following order: Hair, Face Shape, Eyebrows; Eyes; Nose; Mouth; Ears. This order was based on research by Want *et al.* (2003) (described in section 3.3.2), which showed that children aged 5- to 9-years are faster and more accurate in their recognition of outer features alone than for inner features alone (and were fastest for whole faces). Additionally, research by Turner (2004) suggested that working on higher salience features (e.g. as described in Section 3.4.1) early in the construction process led to the highest quality likenesses, and also provided some evidence that simply constraining the order of feature selection was also beneficial.

Instructions to child participants in the composite construction stage of the procedure were compiled with an E-FIT operator who had extensive experience of constructing composites with children. The use of these instructions ensured the E-FIT system was explained in a way which had previously been productive with child witnesses (A. Parry, personal communication, 22nd August, 2003).

The composites were evaluated based on the research regarding the established methods cited in Section 4.3.1. To recap, an important distinction was made between the subjective *likeness* (e.g. using a ranking and/or a rating task) and the objective *utility* (e.g. using a sorting and/or a matching task) of a composite when evaluating composites. Composites constructed in Experiment 5 were evaluated using both of these established measures.

Finally, a ‘typical’ interview for constructing a facial composite image with adults was described in Section 4.2.1. However, when asked about a typical interview for constructing a facial composite image with a child witness, no operators described the same, or even a similar interview procedure to each other. Therefore, there was not enough information to determine a typical interview structure which could be included as a control condition in Experiment 5 for child participants.

8.2 EXPERIMENT 5: COMPOSITE CONSTRUCTION WITH ADULT AND CHILD PARTICIPANTS

8.2.1 METHOD

Design

A between-participant design containing two factors was used. The factor ‘Prompt’ had two levels (Visual and Verbal) and the factor ‘Age’ had four levels (6-, 8-, 10-years and Adult). Participants were pseudo-randomly placed in one of the two prompt interview conditions. Five target videos were used and target was counterbalanced across participants, and within age group with the constraints that the type of prompt and identity of target appeared with equal frequency in each condition and each age group. The age group of each participant interviewed was counterbalanced and the same target never appeared to consecutive participants in order to reduce practice effects of the

operator⁶⁷. In addition, none of the witnesses had prior experience of E-FIT or any other composite construction technique. The experimenter (from now on referred to as the operator) had no exposure to the targets (videos of facial stimuli or identification line-ups). The operator had limited experience of constructing E-FITs (gained from reading the E-FIT manual, conducting approximately 10 E-FITs prior to the experiment and from observing experienced E-FIT operators using the system). Expertise and practice effects were controlled for by counterbalancing the age of participant interviewed. Finally, no participant had constructed a composite before.

The dependent variables were the time taken by participants to select the prompts, participants' selection of the "don't know" prompts and participants' performance at an identification task⁶⁸.

Participants

Sixty children (34 boys and 26 girls) were drawn from a combined school in Buckinghamshire (UK). The children were equally divided into three age groups. The first group comprised 14 boys and six girls from Year 1, aged between 5-years 0-months and 6-years 2-months (mean=5-years 8-months). The second group comprised 12 boys and eight girls from Year 3, aged between 7-years 3-months and 8-years 1-month (mean=7-years 9-months). The third group comprised eight boys and 12 girls from Year 5, aged between 9-years 3-months and 10-years 2-months (mean= 9-years 8-months). Parental and participant consent was gained for each child participant. In addition to the child participants, 20 adults (eight males and 12 females) were drawn from students, staff and visitors to the Open University. They were recruited informally and were not paid for their participation. Their ages ranged from 23-years to 60-years.

⁶⁷ The experimenter did not view the targets at any time prior to or during the Study.

⁶⁸ The likeness and utility of the E-FIT image produced were the dependent variables for Experiments 6 and 7.

No adult or child was paid for his or her participation. All participants spoke English as their first language.

Materials

Facial Stimuli

Five, one minute video sequences were filmed by associates of the operator, using a Sony digital video camera and edited using Adobe Premier 6.5. Each sequence featured a Caucasian adult male aged between late twenties and early thirties who had consented to a video being taken and used in this study. Each target was filmed sitting at a desk and working on a computer. The sequences featured two full face poses of 15 seconds and ten seconds. Each video included a ten second count down screen at the start of the sequence and a blank screen at the end of the sequence to ensure that the operator did not view the targets when showing the stimuli to the participants.

Prompts

The set of visual prompts and the set of verbal prompts described in Study 3 were used with the participants.

Identification Line-ups

A nine-person (one target face along with eight distractor faces) simultaneous line-up was constructed as a PowerPoint Show for each of the five target faces. Each Show included three PowerPoint slides. On the second slide of each Show an image of the target and images of eight other Caucasian males who resembled the target were numbered 1 to 9 in three rows. The first and third slide of each PowerPoint show were blank to ensure that the operator did not view the targets when showing the identification line-ups to the participants. Additionally, each image identification number (1 to 9) was placed in a coloured shape (e.g. the number 1 was inside a red circle, the number 5 was inside a green circle and so on), to further ensure that the

operator could clarify the identification choice made by a child participant without looking at the array (by asking for the number, the shape and the colour of their choice). All images were full face pose with a neutral expression and a blank background. There was no target absent array as the aim of the identification task was to determine whether or not children could recognise the target regardless of their performance with the prompts or on the composite task. Each line-up can be seen in Appendix Vb.

A Sony mini disc portable Walkman (MZ-N910) with a separate Sony microphone (ECM-M5970) was used to record children's responses. A response sheet (the first page of which was based upon the response sheet used with adult participants in Experiment 3, Appendix IVc) was created in order to record participants' selection of prompts. The second page of the response sheet can be seen in Appendix Vc. An instruction sheet was also used to ensure consistent instructions given to children, and to ensure best practice guidelines for interviewing children (Home Office, 2002) were followed as far as possible.

Aspley E-FIT version 3.1a for Windows was used to create composites⁶⁹. The facial stimuli, identification line-ups and E-FIT system were all run on a Dell Inspiron 8000 laptop running Microsoft Windows 2000. A stopwatch was used to ensure a consistent time was spent with all witnesses.

Procedure

Pilot studies were conducted with five participants, consisting of two female adults, two male adults, and a 7-year-old girl. The pilot studies enabled the operator to test the facial stimuli, the identification line-ups, the prompts, the practice interview, the instructions to be used with child and adult participants, displaying the stimuli to a

⁶⁹ For the purpose of Experiment 5, only the White (Caucasian) Male, United Kingdom E-FIT database was available.

participant (without viewing the stimuli) and to test the feasibility of the procedure as a whole.

As a result of the pilot studies the identification array for one of the targets was altered as the line-up was deemed too difficult⁷⁰. A few of the instructions of the E-FIT program ('tagging', moving features and so on) were elaborated to ensure participants understood the available features of the E-FIT program. Pilot data are not included in the results.

In order to add to the ecological validity of the study, the study comprised a number of phases carried out over a two day period (as to resemble those in a criminal investigation). The general procedure followed the instruction sheet for child participants and is summarised below:

Day 1

Participants viewed the video sequence of the relevant target individually in a quiet room outside of the classroom. Participants were instructed to pay close attention to the video, as they would be asked questions about it later. Participants were also informed that the operator would not view the video.

Day 2

All participants were tested individually, in a quiet area free from distractions. Due to school restraints with child participants, a maximum time of just over half an hour was set aside to interview and create a composite with each participant⁷¹.

⁷⁰ No participant correctly selected the target, even when asked to do so immediately after viewing the stimuli video. Therefore a different photograph of the target was used in the array.

⁷¹ The results from the E-FIT operator survey illustrated that average time taken to conduct an interview (from description to composite construction) was just over two hours, with a range from an hour and five minutes to three and a half hours. Therefore, for Experiment 5, time constraints precluded the use of a full interview.

1. Introduction

Each child was introduced to the operator and the purpose of the study was explained. Children were informed that at the end of the interview they would return to their classroom and that, if at any time they wanted to go back to their classroom they could tell the operator. During the introduction it was emphasised to the children that there were no right or wrong answers and that it was acceptable for them to say “I don’t know” to any of the questions they were asked.

2. Practice interview

On the basis of research findings cited in Chapter 2, and guidance from the Home Office (2002), a practice interview was included to further establish rapport and to inform child participants of the stages of the interview, the level of information required of them and to allow them to practice saying “I don’t know” or “I don’t understand (Schwartz-Kenney, Bottoms and Goodman, 1996).

Children were introduced to the prompts (visual or verbal depending on which condition they were in). They were told that the prompts were “<pictures/words> that can be used to answer questions with”. The “not sure” verbal prompt and the ‘?’ visual prompt were explained. Children were then asked to choose a prompt for a feature element relating to their teacher⁷².

3. Target Description Interview

Selection of Prompts

Each child was reminded that the operator had not viewed the target in the video in order to reiterate that the information they were providing was not already known to the

⁷² Consistent with Experiment 4, a child’s teacher was chosen as a face familiar to both the child and the operator but the teacher was not present in the interview room.

operator (Nesbitt and Markham, 1999). Child participants in the verbal prompt interview condition were asked “what is the best *word* for the man’s <insert the name of the feature element>” for all 20 feature elements and 4 configural selections. In the visual prompt condition participants were asked to “select the best *picture* which looks like the man’s <insert the name of the feature element>” for all 20 feature elements and 4 configural selections. Following the selection of the prompts children were asked if there was anything else they wanted to describe about the target. Participants’ selections of prompts and any additional information was noted on the response sheet.

Composite Construction

The operator entered the selection of (visual or verbal) prompts into the E-FIT description boxes in order to create an *initial* face. As described in Study 3, all prompts were designed to map directly onto to the E-FIT description boxes. The operator used a summary form (Appendix Vd) to input participant’s selections of the visual prompts into the E-FIT database. Any responses of “not sure” for the verbal prompts or “?” for the visual prompts were left blank in the description boxes (for which the program selected a default feature). This *initial* face was saved. Before the initial face was shown to the child participant it was explained that they would be shown a black and white face and that the face would not be “an exact photograph of the man” and that it could be changed if they wished to “make it look as much like the man as possible”. When the initial face was shown to the child participant they were asked if it resembled the target face and if there were any changes they would like to make to the face to make it look more like the target. Responses were noted on the response sheet. The operator

explained the main features of the E-FIT system (feature selection, tagging, re-sizing, re-colouring and moving)⁷³.

As described in Section 8.1.2 the seven main facial features were worked through individually in the following order: Hair, Face Shape, Eyebrows; Eyes; Nose; Mouth; Ears. All modifications made to the initial face were noted on the response sheet.

After all of the features had been worked through as much as the child participant wanted to, or the maximum time of 30 minutes had been reached, children were asked if there was anything else they would alter on the composite. Responses to this question were again noted on the response sheet and the E-FIT face was saved again, as a *final* face.

4. Identification task

Once the composite faces had been constructed, children were asked to select the target from the array of 9 faces. Children were told that they were going to be shown some photos and they had to tell the operator “whether there is a picture of the man from the video” and if there was a picture of him could they tell the operator the number next to their choice. The operator also asked for the colour and shape around the number of each child participant’s choice to ensure the correct response made by the child participant was noted (without the operator having to look at the array). Children’s responses were recorded on the response sheet. All children were given fair line-up instructions specifying that the target may or may not be present in the line-up.

⁷³ Artistic enhancement was not used in this study due to time limitations.

5. Discussion and Closure

At the end of the interview, children were thanked, their well-being checked and they were asked if they had any questions about what they had just done. Children were then returned to their classroom and were asked not to discuss the interview. With the consent of the children's parents, all of the interviews were audio taped.

The procedure for adult participants followed the above procedure for child participants with some modifications (e.g. adult participants were asked to think of a friend, rather than their teacher, for the practice task; they were only asked for the number of their choice, rather than the number, shape and colour, in the identification task).

Each interview with a participant lasted just over half an hour consisting of: approximately five minutes for introductions and the practice interview and an average of six minutes selecting the prompts and the remainder of the time spent constructing a composite. At the end of the half an hour time limit participants completed the identification task. Finally, all results were displayed in a 'child friendly' format and sent to each school at the end of the study.

8.2.2 RESULTS

This section considers participants' selection of the prompts in relation to the time taken to select the prompts, participants' use of the "don't know" prompts and participants' performance at the identification task⁷⁴.

⁷⁴ The accuracy of participants' selection of the prompts is not included in the results of Study 4 due to the focus of the study on *composite construction* rather than *descriptions*.

The Effect of Prompts on the Time Taken to Obtain Prompted Descriptions (aim (i))

The time taken by children to select the visual and the verbal prompts was calculated by the use of a stop watch during the interviews and confirmed by listening to and timing the audio recorded interviews. Table 8.1 shows the mean time taken by participants to select the verbal and the visual prompts.

Age	Prompts		Mean (S.D)
	Verbal (S.D)	Visual (S.D)	
6 years	8:21 (1:11)	6:55 (1:16)	7:38 (1:24)
8 years	7:23 (2:15)	5:28 (2:50)	6:26 (1:54)
10 years	6:48 (1:26)	5:18 (1:30)	6:30 (1:37)
Mean (for children)	7:31 (1:44)	5:54 (1:24)	6:43 (1:46)
Adult	4:54 (0:53)	5:34 (0:49)	5:14 (0:54)
Mean (all)	6:51 (1:56)	5:49 (1:17)	6:20 (1:43)

Table 8.1: Mean time taken (minutes:seconds) by participants to select prompts in Experiment 5

Table 8.1 shows that for children, the mean time taken to select both the visual and verbal prompts decreased with age. The mean time taken to select the visual prompts was lower than the mean time taken to select the verbal prompts for children of all ages and the largest difference between these two scores was for 8-year-olds. In contrast, for the adult participants, the mean time taken to select the verbal prompts was lower than the mean time taken to select the visual prompts. Adults were also slightly slower on average than the 8- and 10-year-olds when selecting visual prompts. Finally, the range of mean times taken to select the visual prompts was smaller than the range of mean times taken to select the verbal prompts.

It was hypothesised that there would be a statistically significant effect of *prompt* (i.e. using the visual prompts would lead to descriptions being elicited statistically significantly faster than using the verbal prompts) and a statistically significant effect of *age* (i.e. that descriptions would be elicited statistically significantly faster from older

participants). A 2x4 between participants ANOVA design was employed. The factor Prompt had two levels (Visual and Verbal), and the factor Age had 4 levels (6-, 8-, 10-years and Adult). The dependent variable was the time taken. Results showed a statistically significant main effect of Age [$F_{(3,72)}=11.436$, $p<0.0005$, partial $\eta^2=0.323$] suggesting that older participants were faster selecting the prompts. There was also a statistically significant main effect of Prompt [$F_{(1,72)}=12.444$, $p<0.005$, partial $\eta^2=0.147$], suggesting that descriptions were elicited faster using the visual prompts. These findings support the hypotheses stated above. However, there was also a statistically significant 2-way interaction between Prompt and Age [$F_{(3,71)}= 3.899$, $p<0.05$] illustrated in Figure 8.1.

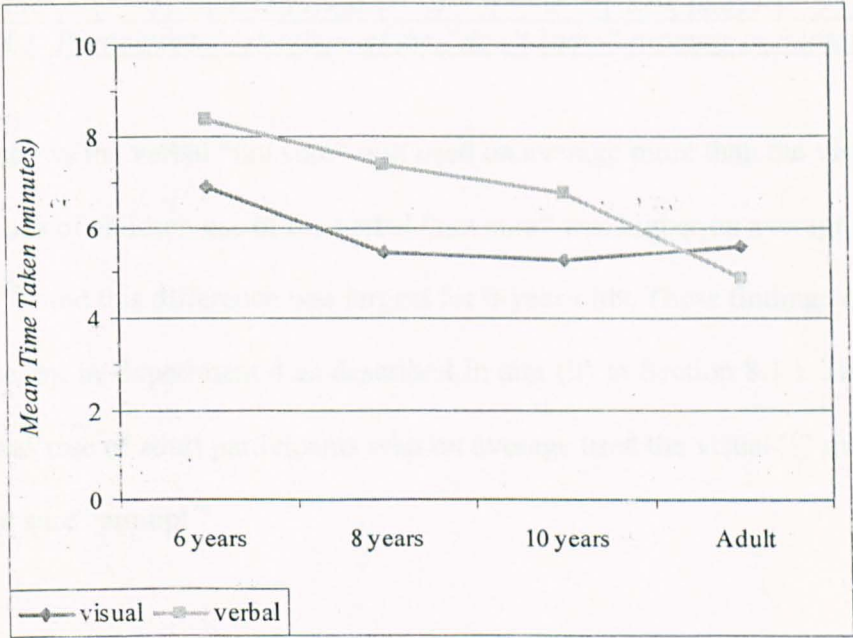


Figure 8.1: Mean time taken by participants to select prompts in Experiment 5

Figure 8.1 shows that the statistically significant interaction appears to be due to the child participants taking longer on average to select the verbal prompts than the visual prompts in contrast to the adult participants. However, post-hoc independent samples t-tests, using the Bonferroni correction for multiple testing⁷⁵, showed the difference

⁷⁵ Adjustments were made to the significance level such that the criterion was set at 0.0125 (0.05/4).

between the time taken to select the visual and verbal prompts only approached significance for 6- and 8-year-olds ($t=2.609$, $df=18$, $p=0.018$; $t=2.644$, $df=18$, $p=0.017$ respectively) and was statistically non-significant for 10-year-olds and adults ($t=2.294$, $df=18$, $p=0.034$; $t=-1.797$, $df=18$, $p=0.089$ respectively).

Participants' Use of the "don't know" Prompts (aim(ii))

Table 8.2 summarises participants' selection of the "not sure" and '?' prompts.

Age	Visual '?'			Verbal "not sure"		
	min	mean (S.D)	max	min	Mean (S.D)	max
6 years	0	1 (1.49)	4	0	3.7 (3.47)	11
8 years	0	1.5 (1.96)	6	0	2.4 (1.78)	6
10 years	0	1.2 (1.93)	6	0	1.4 (1.35)	4
Adult	0	2 (2.13)	5	0	0.8 (0.92)	2
Total	0	1.43 (1.90)	6	0	2.08 (2.31)	11

Table 8.2: Participants' selections of the "don't know" prompts in Experiment 5

Table 8.2 shows the verbal "not sure" was used on average more than the visual '?'. For all age groups of children use of the verbal "not sure" was higher on average than use of the visual '?' and this difference was largest for 6-year-olds. These findings replicate the trends apparent in Experiment 4 as described in aim (ii) in Section 8.1.1. However, the opposite was true of adult participants who on average used the visual '?' more than the verbal "not sure" prompt⁷⁶.

Participants' Performance at an Identification Task Following Composite Construction (aim(iii))

Once the composite faces were constructed using E-FIT, participants were asked if they could select the target they had seen from an array of nine faces. Figure 8.2 shows the percentage of participants who made correct and incorrect identifications (the probability of making a correct identification by chance is 11% and is illustrated by a line in the figure).

⁷⁶ No inferential statistics could be conducted due to the amount of zero uses of the prompts.

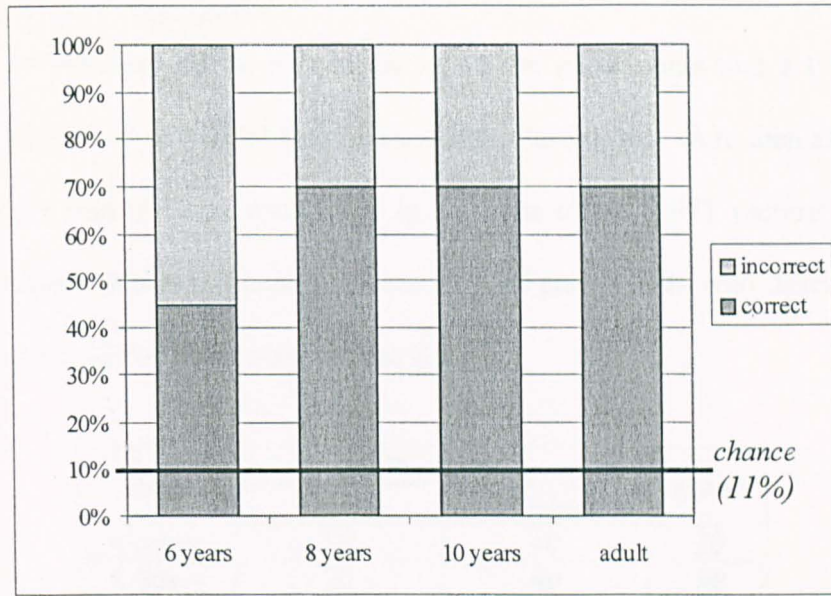


Figure 8.2: Percentage of participants who provided correct and incorrect identifications in Experiment 5

Figure 8.2 shows that the percentage of correct identifications increased from 45% for 6-year-olds to 70% for 8-, 10-year-olds and adults. Even though 6-year-olds provided 25% less correct identifications than all other ages of participants they were still performing at 34% above chance level.

It was hypothesised that the effect of Age would be statistically non-significant. A Chi-Square test showed that the relationship between overall identification result and age of participant was statistically non-significant [$X^2=0.255$, $df=2$, $p=0.267$], supporting the hypothesis stated above and replicating the findings of Experiment 4. Finally, the relationship between overall identification result and prompt condition was also statistically non-significant (visual or verbal) [$X^2=0.487$, $df=1$, $p=0.321$].

Participants' performance at the identification task in relation to the likeness and utility scores for their composites are considered in the end of the result section for Experiment 7 (Section 8.4.2).

Additional E-FIT changes

At the end of the interview it was explained to the participants that a typical police interview could go on as long as they wanted it to. Participants were then asked if there were any other changes they would like to be made to the E-FIT picture if there had been more time. Table 8.3 shows the percentage of participants who described further changes they would like to see made to the E-FITs.

Age	Prompts		Mean
	Visual	Verbal	
6 years	10	40	20
8 years	20	40	30
10 years	90	90	90
Adult	100	80	90

Table 8.3: Percentage of participants who stated further changes to be made to the E-FIT in Experiment 5

It can be seen from Table 8.3 that almost all 10-year-olds and adults listed at least one further change that they would like to see made to the E-FIT they were working on if they had more time. In contrast, only around a quarter of 6 and 8-year-olds listed further changes to be made to the E-FIT if they had more time. For 6 and 8-year-olds a higher percentage of children who used verbal prompts stated further changes should be made compared to those who used visual prompts. These figures were then same for the 10-year-olds and the reverse was true for adult participants. A Chi-Square test showed that there was statistically significant relationship between the percentage of participants who described further changes they would like to see made to the E-FITs and the age of participant [$X^2=34.987$, $df=3$, $p<0.0005$].

Figure 8.3 shows the mean number of changes participants listed they would like to be made (if the interview was longer)⁷⁷.

⁷⁷ Changes were coded as follows: if a participant mentioned a feature (e.g. “change eyes”) this counted as 1 change. If a participant mentions a feature and states the particular change to be made (e.g. “make eyes bigger”) this counted as 2 changes (i.e. one for “eyes” and one for “bigger”).

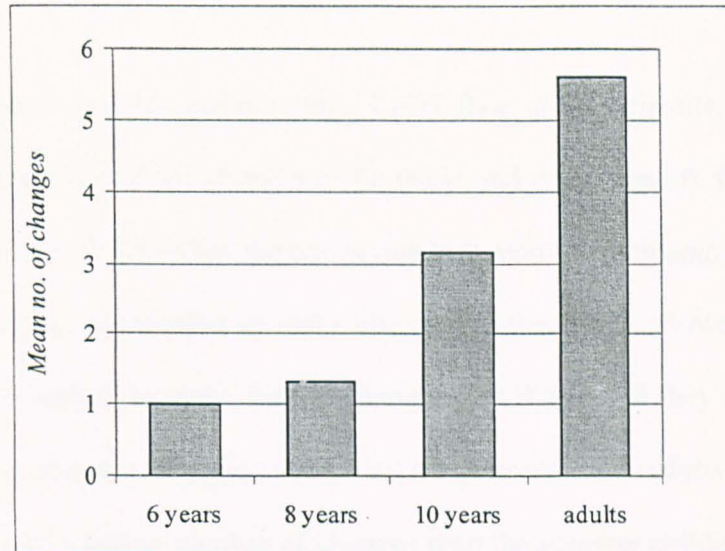


Figure 8.3: Mean number of changes listed by participants to be made to their E-FIT in Experiment 5

From Figure 8.3 it can be seen that the mean number of changes to be made to the final E-FIT image increased with age⁷⁸. In summary, very few 6- and 8-year olds wanted to make further changes and if they did they only wanted to make an average of one change. However, almost all 10-year-olds and adults wanted to make changes.

8.2.3 DISCUSSION OF EXPERIMENT 5 RESULTS

Are children and adults able to construct computerised composites using the visual and verbal prompts in an experimental setting? (aim (iv))

The findings illustrate that children from the age of 6-years can produce facial composites of an unfamiliar face using the visual and verbal prompts and the computerised composite system E-FIT. All participants selected a set of visual or verbal prompts in order to provide a description of the target and went on to construct a composite within approximately half an hour. All the composites constructed were of a white male, within the correct age range for the target. An image of each composite constructed can be seen in Appendix Ve.

⁷⁸ An ANOVA could not be employed due to the uneven variances within the age groups.

When asked about the quality of the *initial* E-FIT face, all participants, including all age groups of children, asked for changes to be made and could specify the changes to be made. Furthermore, at the end of the composite task many participants, including all age groups of children, still wanted to make changes to the E-FIT. However, very few 6- and 8-year olds wanted to make further changes and if they did they only specified an average of one change. Whereas, almost all 10-year-olds and adults wanted to make changes and listed a higher number of changes than the younger children. If participants had been provided with unlimited time to construct a composite it is probable that adults and older children would have spent longer and made further changes to composites. Therefore, it is unlikely that adults and 10-year olds optimum performance with the E-FIT system was being measured. However, the small number of younger children wanting to make further changes, and the low number of changes listed by those who did want to make changes, suggests that the composites obtained from these younger age groups were probably nearer completion and that allowing more time would not have resulted in significantly different composites. These findings suggest support for the *quantitative* change in the development facial processing described in Section 3.2.2 i.e. younger children did not want to make as many changes to the composite they were presented with.

However, two factors complicate this issue and prevent any definite conclusions being drawn with regards to the relationship between the composites actually produced and those which may have been produced if unlimited time were allowed.

First, due to children's language development (as described in Section 2.1.3) younger children may not have fully understood the question regarding additional changes, or misunderstood it. Or younger children may have been unable to determine proactively what they might do should they be allowed more time whereas adults were more aware of the further changes that could be made within the E-FIT program. Some evidence for this is provided by the fact that none of the children who responded that they would not make any additional changes, also indicated to the researcher that the composite was completed prior to being asked. Rather than inform the researcher that the composite could not be further improved without prompting, they were merely responding to the question regarding additional changes, suggesting they may well have continued working on the image if allowed. Likewise the older children and adults may also have responded inaccurately, either by under- or over-estimating the changes they would make. Perhaps an appropriate interview technique similar to the prompts used during the interview process could be used with children during composite construction.

Secondly, it is far from clear that the final composite produced when unlimited time is allowed is the optimum image. There is some evidence that after a while some participants actually make changes that introduce errors and make the composite less, not more, like the target (Pike, Brace, Kynan and Turner, 2004). Thus it cannot be assumed that the composites that the older children and adults may have gone on to produce would have actually been more accurate than those produced at the end of the time limit. Although the exact implications of imposing a time limit are therefore impossible to draw, the data pertaining to additional changes do show that the final composite produced in this study is not indicative of the composite that would be produced if no time limits were in place.

The holistic system E-FIT also allows witnesses to make configural changes to a composite. Children and adults may use both kinds of facial information (featural and configural) depending on the requirements of the task and its degree of difficulty. Therefore working on a feature in E-FIT, although shown within a whole face, may disrupt configural processing abilities which are weaker in young children (especially 6-years of age and under) compared with older children and adults (e.g. Friere and Lee, 2001; Mondloch *et al.*, 2002). Adults and older children may be able to utilise the holistic aspect of E-FIT more than younger children and switch between the two. They may also want to go on to make more configural changes than younger children, resulting in better quality composites.

The Effect of Prompts on the Time Taken to Obtain Prompted Descriptions (aim (i))

The hypotheses based on the findings from Experiment 4 were supported: using the visual prompts led to descriptions being elicited statistically significantly faster than when using the verbal prompts, and descriptions were elicited from older participants significantly faster than from younger participants. However, post-hoc examinations of the data showed that the difference between the time taken to select the visual and verbal prompts only approached significance for the 6- and 8-year-olds. Interestingly, adult participants were slightly slower on average than the 8- and 10-year olds when selecting the visual prompts. This difference cannot simply be due to adults' expertise with reading as will be explained by adults' use of the "don't know" prompts below.

Participants' Use of "don't know" Prompts (aim (ii))

Results replicated the findings from Experiment 4 for child participants: the verbal "not sure" was used on average more than the visual "?". However, the opposite was true of adult participants who on average used the visual "?" more than the verbal "not sure"

prompt. Additionally, on average adults selected the visual '?' more than 6-, 8- and 10-year-olds.

As described above, adult participants were spending slightly longer on average than the 8- and 10-year olds when selecting the visual prompts. This, taken together with adults' use of the "don't know" prompts, suggests that the differences observed with adult participants cannot simply be due to their reading expertise. Perhaps the amount of implicit information contained in the visual prompts (compared to the verbal prompts) may have encouraged the adult participants to study them in more detail, therefore requiring a longer length of time and perhaps resulting in a higher use of the '?' option. Younger children, in contrast, may not have spent time analysing the visual prompts beyond their explicit representation and may instead have required more attention to read the verbal prompts and perhaps interpret their meaning. In other words, the verbal prompts were more generic for adults and the visual prompts were more generic for children. Anecdotal support can be provided for such an explanation from adults' comments to the prompts gathered in Study 3. Adults who preferred the verbal prompts stated that "I can describe things better with words" and "it's easier to describe using the words, you have more ability to qualify them" comments were also made that "the pictures [the visual prompts] showed more detail".

The mean number of uses of both "not sure" and '?' were higher for child participants in the present study than in Experiment 4. Previous research has shown that children do not believe that they can give information which will aid a criminal investigation as they may believe that an adult knows everything about an event, and that an adult knows the answers to the questions they are asking (Nesbitt and Markham, 1999; Saywitz, Nathanson, Snyder, 1993). Therefore, the fact that in the present study the

operator did not view the target with participants whereas in Experiment 4 child participants were aware that the experimenter viewed the target with them may have resulted in the increased use.

As described in the discussion of Study 3 (Section 7.4.1), of particular importance was the finding that children used the “don’t know” prompts, due to concerns which have been raised about option posing forms of questions (e.g. Poole and White, 1993). Importantly, in both Study 3 and Study 4 children appeared to be comfortable selecting both visual and verbal “don’t know” prompts (as illustrated by their use of these prompts) rather than attempting to answer questions which they did not fully understand (Saywitz and Snyder, 1993) or could not remember the answers to. Children’s use of the prompts also provides support for the inclusion of a practice task in the use of these prompts (Schwartz-Kenney, Bottoms and Goodman, 1996).

Participants’ Performance at an Identification Task following Composite Construction (aim (iii))

An identification task was included in the present studies to determine whether children had attended to the target. Results showed that all groups of participants performed at above chance level in the identification task following composite construction. The hypothesis that there would not be a statistically significant difference between the different age groups was supported, replicating the findings from Experiment 4 and the meta analysis conducted by Pozzulo and Lindsay (1998). Although not statistically significant, correct identifications increased with age (as they did in Experiment 4). Specifically, correct identifications increased between 8- and 10-year-olds in Experiment 4 whereas the findings from the present study showed 8- and 10-year-olds performing at the same level as adults. This supports the meta analysis research findings of Pozzulo and Lindsay, who found that children reached an adult correct identification

rate at a much earlier age (around 9- to 10-years) than previously proposed by Chance and Goldstein (1984).

Although all participants performed at above chance level, there did appear to be a group of children who had simply failed to attend to the target (55% of 6-year-olds made an incorrect identification). It is possible that this group of children did not encode the target face. This possibility will be discussed at the end of Section 8.4.3 where participants' performance at the identification task will be considered in relation to the quality of their composites.

8.3 EXPERIMENT 6 – SUBJECTIVE COMPOSITE EVALUATION (RANKING AND RATING TASKS)

8.3.1 METHOD

Design

A within-participants design containing two factors was used. The factor 'Prompt' had two levels (Visual and Verbal). The factor 'Age' had four levels (6-, 8- and 10-years and Adult). The same participants undertook both the ranking and the rating task. In both tasks the participants viewed all of the composites for two of the five targets, making a total of 32 composites viewed per participant. The dependent variables were the ranking (from best to worst) of the 16 composites for each target face and the likeness rating (on a scale of 1-10). In addition to this, participants were asked to provide a measure of "distinctiveness" of the target faces viewed (on a scale of 1-10).

The composites were labelled as A to P within each set, with the position of images from each condition (Prompt and Age) randomised. The composites were also rearranged between participants.

Participants

Twenty-five adults (ten males and 15 females) were drawn from students, staff and visitors to The Open University. They were recruited informally and were not paid for their participation. Their ages ranged from 23 years to 61 years. All participants spoke English as their first language. In addition, none of the participants had taken part in the construction of the composites (Experiment 5).

Materials

Three images of each target face were captured from each one-minute video sequence using Windows Media Player v9 and Jasc PaintShop Pro 6.0. The three images of each target included different perspective views, though a full face view was always presented, to provide participants with as much information about the target's facial features as possible. Images were printed on an A4 sheet as a portrait colour print measuring 6cm by 8cm. Figure 8.4 shows the three images for Target 5. Images for the all of the targets can be seen in Appendix Vf.



Figure 8.4: Facial images of target 5 for ranking and rating tasks

The 80 composites (constructed in Experiment 5) were printed in greyscale at 72 d.p.i. resolution (E-FIT's maximum output). An image of every composite constructed can be seen in Appendix Ve. For the purposes of the ranking and rating tasks each composite

was printed on an A4 sheet, and measured approximately 15cm by 10cm. The composites were labelled as A to P within each set.

An instruction sheet was provided along with brief background information about facial composites (including an example of a photograph and corresponding E-FIT created by an experienced E-FIT operator). This was to inform participants that the aim of an E-FIT is to produce a “likeness” of a suspect (see Appendix Vg). Finally, a participant response sheet, in the form of a table with columns for distinctiveness (1-10), rank position of each composite (1-4) and score for each composite (1-10) (see Appendix Vh) was provided.

Procedure

Participants were given an instruction sheet and a response sheet in the form of a table (as described in the ‘Materials’ section above). The general procedure followed the instruction sheet and is summarised below:

Participants were instructed that they were about to see two sets of face images, labelled A to P, and that each set would consist of three photographs of a man and 16 E-FITs (facial composite images) that had been made of the man depicted in the photographs.

It was stressed to participants that the labels (A to P) were simply identifying letters to allow them to record responses and that they were allocated entirely randomly and did not give any indication as to a particular order. Participants were also informed that there was no time limit and that there were no right or wrong answers. Participants were then asked to provide a number of scores:

(1) Distinctiveness of Target

Firstly, they were instructed to examine the photographs of the target and decide how *distinctive* the person looked on a scale of 1 (very average looking) to 10 (very distinctive). Participants who asked for a definition of distinctiveness were told that a distinctive person would “stand out in a crowd”.

(2) Ranking of Composites

Participants were then asked to compare and sort the E-FITs (A to P) *equally* into 4 ordered groups as follows:

Group 1 Best (4 E-FITs) look <i>most</i> like the photo	Group 2 Next Best (4 E-FITs)	Group 3 Next Worst (4 E-FITs)	Group 4 Worst (4 E-FITs) look <i>least</i> like the photo
-----------------------------------------------------------------------------------	--------------------------------------------------	---------------------------------------------------	-------------------------------------------------------------------------------------

The group number for each E-FIT was recorded in the column marked “Group Number” (i.e. Group 1, Group 2, Group 3 or Group 4).

(3) Rating of Composites

Participants were asked to provide each E-FIT (A to P) with a score from 1 to 10 (where 1 is “nothing like the person” and 10 is “exactly like the person”) in the column marked ‘Rating’. Participants were told that they could provide the same rating score for more than one E-FIT.

This whole procedure was repeated for a second target using a new response sheet. At the end of the task, participants were thanked and de-briefed.

8.3.2 RESULTS

This section considers data for the ranking and ratings tasks. Due to the focus on composites, the data for each task was analysed first by-item (by composite), where the

composites are treated as a random effect and the data was collapsed (averaged) across participants. However, this analysis does not take account of the variability of the participants. To determine if the effect would generalise from the sample to the population the data was then analysed by-participant, where participants were treated as a random effect and the data was collapsed across composites. Finally, the results from the by-item and by-participant analyses were combined.

Effects of Target on the Data

Table 8.4 shows a breakdown of the distinctiveness ratings provided by participants for each target⁷⁹.

Target Face	Mean (S.D.)	Min.	Max.
1	5.30 (1.70)	3	7
2	5.22 (1.99)	2	7
3	5.50 (2.37)	1	8
4	6.44 (1.24)	4	8
5	5.44 (1.67)	3	8

Table 8.4: Distinctiveness ratings provided by participants for each target face used in Experiment 5

Table 8.4 illustrates that the mean and standard deviation scores (S.D.) for the faces were fairly comparable in terms of distinctiveness. Target 4 received the highest mean distinctiveness score (6.44). The range of scores given for all targets is quite high. A 1-way between participants ANOVA design was employed. The factor Target had five levels (Target 1 to Target 5). Results showed a statistically non-significant main effect of Target [$F_{(4,42)}=0.655$, $p=0.627$] on distinctiveness rating.

Additionally, the effects of Target on the ranking and rating of E-FITs were tested for. Results showed Target effects were equal in all conditions (Age and Prompt) and are

⁷⁹ *n* for targets 2, 4, and 5 is 9 (rather than 10), as distinctiveness ratings were provided by all but one participant.

not considered further in this results section. Full ANOVA calculations are provided in Appendix Vi.

Ranking Data

Ranking Data (considered by-item)

Note that for rankings a score of 1 signifies the best ranking and a score of 4 the worst ranking. Table 8.5 summarises the mean ranks when considered by-item (collapsed across participants).

Age	Mean (<i>S.D</i>)	Prompt	Mean (<i>S.D</i>)
6 years	3.08 (0.72)	Visual	2.83 (0.70)
		Verbal	3.34 (0.69)
8 years	2.73 (0.64)	Visual	2.67 (0.62)
		Verbal	2.78 (0.69)
10 years	2.63 (0.72)	Visual	2.85 (0.71)
		Verbal	2.41 (0.69)
Adult	1.86 (0.57)	Visual	1.81 (0.68)
		Verbal	1.91 (0.46)
	<i>Total</i>	<i>Visual</i>	2.54 (0.78)
		<i>Verbal</i>	2.61 (0.81)

Table 8.5: Mean ranks of composites (by-item analyses) in Experiment 6

From Table 8.5 it can be seen that overall, the mean rank awarded to composites decreased (indicating a better rank) with the age of participants who produced the composites. Overall, the mean rank awarded to composites produced using visual prompts was lower (i.e. better) than the mean rank for the composites created using verbal prompts. Finally, the mean rank awarded to composites created by 6- and 8-year-olds and adults using visual prompts were lower (i.e. better) than the mean rank for composites created using verbal prompts for these age groups. However for 10-year-olds the reverse was true.

Table 8.6 shows the percentage of composites ranked within each group when considered by-item (collapsed across participants).

Rank Group	vs ⁸⁰	vb	6-years			8-years			10-years			Adult		
				vs	vb		vs	vb		vs	vb		vs	vb
1 (best)	25.8	24.3	8	12	4	15	11	19	26	27	25	51	53	49
2	23	27	17.5	24	11	27.5	21	34	24.5	23	26	30.5	24	37
3	28.8	21.3	25.5	31	20	32.5	37	28	27	29	25	15	18	12
4 (worst)	22.5	27.5	49	33	65	25	31	19	22.5	21	24	3.5	5	2

Table 8.6: Percentage of composites within each rank group (by-item analyses) in Experiment 6

Table 8.6 shows that overall, a higher percentage of the composites created by adults received a ‘best’ rank than composites created by children and a lower percentage of composites created by adults received a ‘worst’ rank than composites created by children. Although the data shows a consistent increase in the rating of composites with age, composites produced by all ages of participants were ranked within *all* groups (best to worst): over 25% of composites created by 6-year olds were ranked within the top two groups; this figure was almost 50% for composites created by 8-year-olds and over 50% for composites created by 10-year-olds.

Overall, a higher percentage of the composites created using visual prompts received a ‘best’ rank than composites created using verbal prompts and a lower percentage of composites created using visual prompts received a ‘worst’ rank than composites created using verbal prompts. Finally, composites produced by using both sets of prompts were ranked within *all* groups (best to worst):

It was hypothesised that there would be a general improvement in the likeness of composites produced with age. A 4x2x2 between participants ANOVA design was

⁸⁰ vs = visual, vb = verbal.

employed. The factor Age had 4 levels (6-, 8-, 10-years and Adult), the factor Prompt had 2 levels (Visual and Verbal), and the factor Identification had 2 levels (Correct and Incorrect)⁸¹. The dependent variable was the mean rankings.

Results showed the main effect of Prompt was statistically non-significant [$F_{(1,64)}=0.466$, $p=0.497$] and the 2-way Age by Prompt interaction was also statistically non-significant [$F_{(3,64)}=0.865$, $p=0.464$]. There was a statistically significant main effect of Age [$F_{(3,64)}=11.289$, $p<0.0005$, partial $\eta^2=0.346$] suggesting that older participants produced composites that were given higher rankings, thus supporting the hypothesis stated above. This effect was analysed further using a post-hoc Scheffé test, the results of which are shown in Table 8.7.

Age 1	Age 2	Mean Difference	Significance
6-years	8-years	0.480	$p=0.234$
	10-years	0.695	$p<0.05$
	Adult	1.445	$p<0.0005$
8-years	6-years	-0.480	$p=0.234$
	10-years	0.215	$p=0.830$
	Adult	0.965	$p<0.005$
10-years	6-years	-0.695	$p<0.05$
	8-years	-0.215	$p=0.830$
	Adult	0.750	$p<0.05$
Adult	6-years	-1.445	$p<0.0005$
	8-years	-0.965	$p<0.005$
	10-years	-0.750	$p<0.05$

Table 8.7: Post-hoc Scheffé test of the effect of Age on the ranking scores in Experiment 6

The results shown in Table 8.7, together with the data described above, suggest that the composites constructed by adult participants were ranked as statistically significantly better likenesses than the composites constructed by children and the composites constructed by 10-year-olds were ranked as statistically significantly better likenesses than the composites constructed by 6-year-olds. Interestingly, the difference between

⁸¹ The between participants factor Identification will be considered at the end of Section 8.4.2.

the mean ranking scores for composites produced by 6- and 8-year-olds and 8- and 10-year-old children was statistically non-significant.

Ranking Data (considered by-participant)

Table 8.8 summarises the mean ranks when considered by-participant (collapsed across items).

Age	Mean (<i>S.D</i>)	Prompt	Mean (<i>S.D</i>)
6-years	3.16 (0.25)	Visual	2.88 (0.50)
		Verbal	3.43 (0.32)
8-years	2.67 (0.29)	Visual	2.68 (0.40)
		Verbal	2.65 (0.42)
10-years	2.46 (0.41)	Visual	2.61 (0.68)
		Verbal	2.31 (0.45)
Adult	1.71 (0.32)	Visual	1.57 (0.45)
		Verbal	1.85 (0.35)
	<i>Total</i>	<i>Visual</i>	2.44 (0.16)
		<i>Verbal</i>	2.56 (0.16)

Table 8.8: Mean ranks of composite (by-participant analyses) in Experiment 6

From Table 8.8 it can be seen that the pattern of results by age and prompt were the same as the by-item data (Table 8.5). The mean rank awarded to composites created by 6-year-olds and adults using visual prompts were lower (i.e. better) than the mean rank for composites created using verbal prompts for these age groups while the opposite was true for both 8- and 10-year-olds.

A 4x2 within-participants ANOVA design was employed⁸². The factor Age had 4 levels (6-, 8- and 10-years and Adult), and the factor Prompt had 2 levels (Visual and Verbal). The dependent variable was the mean rankings. The ANOVA revealed the same pattern of results as that conducted above by-item i.e. a statistically significant main effect of Age [$F_{(3,72)}=65.538$, $p<0.0005$, partial $\eta^2=0.732$] (again supporting the hypothesis that there would be a general improvement in the likeness of composites produced with age)

⁸² The between participants factor Identification could not be included in the by-participant data.

and a statistically non-significant main effect of Prompt [$F_{(1,24)}=3.947$, $p=0.058$]). However, there was one exception in that the 2-way interaction between Age and Prompt for the by-participants analyses was statistically significant [$F_{(2.095,50.272)}=7.902$, $p<0.005$, partial $\eta^2=0.248$]⁸³.

The statistically significant Age by Prompt interaction when analysed by-participant but not when analysed by-item can be explained by the fact there is a large variation around the mean scores for Age and Prompt when analysed by-item (see the large standard deviations in Table 8.5) and only a small variation around mean scores for Age and Prompt when analysed by-participant (see the small standard deviations in Table 8.8).

Post-hoc paired-samples t-tests, using the Bonferroni correction for multiple testing⁸⁴, showed the ranking score for composites constructed by 6-year-olds using the visual prompts was statistically significantly higher than the ranking score for composites constructed by 6-year-olds using the verbal prompts ($t=4.109$, $df=24$, $p<0.0125$). The same was true for composites constructed by adults ($t=2.280$, $df=24$, $p<0.0125$). However, the difference between the ranking score for composites constructed using the visual and verbal prompts for 8- and 10-year-olds was statistically non-significant ($t=-0.257$, $df=24$, $p=0.799$; $t=-1.852$, $df=24$, $p=0.076$ respectively).

Using the results from the by-item and by-participant analyses min F' was calculated for the effects. Min F' combines the different F ratios calculated above to form a single value. The formula used to calculate Min F' was: $Min F' = F1F2 / (F1+F2)$ (Clarke, 1973). The degrees of freedom were also calculated for the effects. The formula used to

⁸³ Mauchly's test of sphericity was statistically significant [$W=0.498$, $df=5$, $p<0.05$], so the Greenhouse-Geisser epsilon correction was applied to the degrees of freedom and subsequent p-value.

⁸⁴ Adjustments were made to the significance level such that the criterion was set at 0.0125 (0.05/4).

calculate the degrees of freedom was: $df = (F1F2)^2 / (F1^2 / n2 + F2^2 / n1)$ (Clarke, 1973).

The results of the Min F' and degrees of freedom calculations show a similar pattern of results as the previous by-item calculations i.e. a statistically significant main effect of Age [$F_{(3, 100.401)}=9.6301$, $p<0.05$] (supporting the hypothesis that there would be a general improvement in the likeness of composites produced with age), a statistically non-significant main effect of Prompt [$F_{(1,95.735)}=0.416$, $p>0.05$], and a statistically non-significant 2-way Age by Prompt interaction [$F_{(3,88.692)}=0.438$, $p>0.05$]. Therefore, the statistically significant main effect of Age can be assumed to generalise to the whole population and to a different set of composites (i.e. it is not that there is a particular participant or composite which is affecting the results unduly). The statistically significant 2-way Age by Prompt interaction for the by-participant analyses cannot be generalised.

Rating Data

Rating Data (considered by-item)

Note that for ratings a score of 1 signifies the lowest rating and a score of 10 the highest rating (ie. opposite to ranking). Table 8.9 summarises the mean ratings when considered by-item (collapsed across participants).

Age	Mean (<i>S.D</i>)	Prompt	Mean (<i>S.D</i>)
6 years	2.95 (1.55)	Visual	3.46 (1.43)
		Verbal	2.43 (1.55)
8 years	3.99 (1.44)	Visual	4.08 (1.55)
		Verbal	3.91 (1.40)
10 years	4.45 (1.73)	Visual	4.04 (1.89)
		Verbal	4.82 (1.54)
Adult	6.29 (1.33)	Visual	6.69 (1.43)
		Verbal	5.89 (1.15)
	<i>Total</i>	<i>Visual</i>	4.57 (1.98)
		<i>Verbal</i>	4.27 (1.88)

Table 8.9: Mean ratings of composites (by-item analyses) in Experiment 6

From Table 8.9 it can be seen that overall, the mean rating awarded to composites increased with the age of participants who produced the composites. However, the mean rating awarded to composites constructed by 6-year-olds was almost 30% and this figure rose to almost 45% for 10-year-olds. The mean rating awarded to composites produced using visual prompts was higher than the mean rating for the composites created using verbal prompts. Finally, the mean rating awarded to composites created by 6-, 8- year-olds and adults using visual prompts were higher than the mean rating for composites created using verbal prompts for these age groups. However for 10-year-olds the reverse was true. The highest standard deviation score was for the rating of composites created by 10 year olds using visual prompts. This was due to two composites being consistently rated very low on average.

Table 8.10 shows the percentage of composites receiving each rating score when considered by-item (collapsed across participants).

Rating	vs ⁸⁵	vb	6-years			8-years			10-years			Adult		
				vs	vb		vs	vb		vs	vb		vs	vb
1 (low)	14.5	18	35.5	22	49	15	16	14	13	18	8	1.5	2	1
2	13	12.3	19	19	19	14.5	13	16	13.5	18	9	3.5	2	5
3	12.5	11.5	10.5	15	6	17	17	17	13	11	15	7.5	7	8
4	10.3	12.8	13	14	12	12	9	15	12	10	14	9	8	10
5	11.5	11.8	7	11	3	14.5	15	14	12	11	13	13	9	17
6	10.5	11.8	6	8	4	12.5	14	11	12	12	12	14	8	20
7	13	10.5	6	7	5	7.5	9	6	13	10	16	20.5	26	15
8	8.8	8	2.5	3	2	6.5	6	7	9	10	8	15.5	16	15
9	3.8	2.8	0.5	1	0	0.5	1	0	2.5	0	5	9.5	13	6
10 (high)	2.3	0.8	0	0	0	0	0	0	0	0	0	6	9	3

Table 8.10: Percentage of composite receiving each rating score (by-item analyses) in Experiment 6

Table 8.10 shows that overall, a larger percentage of the composites created by adults received a high rating than composites created by children. None of the children's composites received a top rating of 10. Additionally, a lower percentage of composites created by adults received a low rating than composites created by children. Although the data shows a consistent increase in the rating of composites with age, around 15% of composites constructed by 6 year olds received a rating of 6 and over, this figure was over 25% for composites constructed by 8-year-olds and over 35% for composites constructed by 10-year-olds.

A higher percentage of the composites created using visual prompts received a 'high' rating than composites created using verbal prompts and a lower percentage of composites created using visual prompts received a 'low' rating than composites created

⁸⁵ vs = visual, vb = verbal.

using verbal prompts. Composites produced regardless of type of prompt received the whole range of scores (from 1 to 10).

It was hypothesised that there would be a general improvement in the likeness of composites produced with age. A 4x2x2 between participants ANOVA design was employed. The factor Age had 4 levels (6-, 8- and 10-years and Adult), the factor Prompt had 2 levels (Visual and Verbal), and the factor Identification had 2 levels (Correct and Incorrect)⁸⁶. The dependent variable was the mean ratings. Results showed the main effect of Prompt was statistically non-significant [$F_{(1,64)}=0.422$, $p=0.519$] and the 2-way Age by Prompt interaction was also statistically non-significant [$F_{(3,64)}=1.005$, $p=0.396$]. There was a statistically significant main effect of Age [$F_{(3,64)}=12.394$, $p<0.0005$, partial $\eta^2=0.367$] suggesting that older participants produced composites that were given higher ratings and supporting the hypothesis stated above. This effect was analysed further using a post hoc Scheffé test, the results of which are shown in Table 8.11.

Age 1	Age 2	Mean Difference	Significance
6-years	8-years	-1.050	$p = 0.231$
	10-years	-1.500	$p < 0.05$
	Adult	-3.345	$p < 0.0005$
8-years	6-years	1.050	$p = 0.231$
	10-years	-0.450	$p = 0.847$
	Adult	-2.295	$p < 0.0005$
10-years	6-years	1.500	$p < 0.05$
	8-years	0.450	$p = 0.847$
	Adult	-1.845	$p < 0.005$
Adult	6-years	3.345	$p < 0.0005$
	8-years	2.295	$p < 0.0005$
	10-years	1.845	$p < 0.005$

Table 8.11: Post hoc Scheffé test of the effect of Age on the rating scores in Experiment 6

⁸⁶ The between participants factor Identification will be considered at the end of Section 8.4.2.

The results shown in Table 8.11, together with the results described above, suggest that the composites constructed by adult participants were rated as statistically significantly better likenesses than the composites constructed by children and the composites constructed by 10-year-olds were rated as statistically significantly better likenesses than the composites constructed by 6-year-olds. Interestingly, the difference between the mean rating scores for composites produced by 6- and 8-year-olds and 8- and 10-year-old children was statistically non-significant (i.e. mirroring the ranking data results described previously).

Rating Data (considered by-participant)

Table 8.12 summarises the mean rating score when considered by-participant (collapsed across items).

Age	Mean (<i>S.D</i>)	Prompt	Mean (<i>S.D</i>)
6-years	2.95 (0.80)	Visual	3.46 (1.18)
		Verbal	2.43 (0.83)
8-years	4.00 (0.86)	Visual	4.08 (1.23)
		Verbal	3.91 (0.88)
10-years	4.45 (0.97)	Visual	4.04 (1.35)
		Verbal	4.85 (1.27)
Adult	6.30 (1.05)	Visual	6.69 (1.27)
		Verbal	5.90 (1.15)
	Total	Visual	4.57 (0.65)
		Verbal	4.27 (0.71)

Table 8.12: Mean ratings of composite (by-participant analyses) in Experiment 6

From Table 8.12 it can be seen that the pattern of results by age and prompt is the same as the by-item data (Table 8.9).

A 4x2 within-participants ANOVA design was employed to analyse these data. The factor Age had 4 levels (6-, 8- and 10-years and Adult), and the factor Prompt had 2 levels (Visual and Verbal)⁸⁷. The dependent variable was the mean ratings. The

⁸⁷ The between participants factor Identification could not be included in the by-participant data

ANOVA revealed the same statistically significant main effect of Age [$F_{(3,72)}=72.011$, $p<0.0005$, partial $\eta^2=0.750$] (again supporting the hypothesis that there would be a general improvement in the likeness of composites produced with age). However, the by-participant analysis also revealed a statistically significant main effect of Prompt [$F_{(1,24)}=5.163$, $p<0.05$, partial $\eta^2=0.177$] and a statistically significant interaction between Age and Prompt [$F_{(3,72)}=8.444$, $p<0.0005$, partial $\eta^2=0.260$].

The statistically significant main effect of Prompt and the statistically significant 2-way Age by Prompt interaction when analysed by-participant but not when analysed by-item can be explained by the fact there is a large variation around the mean scores for Age and Prompt when analysed by-item (see the large standard deviations in Table 8.9) and only a small variation around mean scores for Age and Prompt when analysed by-participant (see the small standard deviations in Table 8.12).

Post-hoc paired-samples t-tests, using the Bonferroni correction for multiple testing⁸⁸, showed the rating score for composites constructed by 6-year-olds using the visual prompts was statistically significantly higher than the rating score for composites constructed by 6-year-olds using the verbal prompts ($t=-4.111$, $df=24$, $p<0.0125$). The same was true for composites constructed by adults ($t=-3.296$, $df=24$, $p<0.0125$). In contrast, the rating score for composites constructed by 10-year-olds using verbal prompts was statistically significantly higher than the rating score for composites constructed by 10-year-olds using the visual prompts ($t=2.310$, $df=24$, $p<0.0125$). Finally, the difference between the rating score for composites constructed using the visual and verbal prompts for 8-year-olds was statistically non-significant ($t=-0.674$, $df=24$, $p=0.507$).

⁸⁸ Adjustments were made to the significance level such that the criterion was set at 0.0125 (0.05/4).

Using the results from the by-item and by-participant analyses min F' was calculated for the effects. The results of the Min F' and degrees of freedom calculations show a similar pattern of results as the previous calculations for the by-item analysis i.e. a statistically significant main effect of Age [$F_{(3,100.391)}=10.547$, $p<0.05$] (supporting the hypothesis that there would be a general improvement in the likeness of composites produced with age), a statistically non-significant main effect of Prompt [$F_{(1,91.652)}=0.3907$, $p>0.05$] and a statistically non-significant 2-way Age by Prompt interaction [$F_{(3,87.067)}=0.378$, $p>0.05$]. Therefore, the statistically significant main effect of Age can be assumed to generalise to the whole population and to a different set of items (composites) (i.e. it is not that there is a particular composite which is affecting the results). The statistically significant main effect of Prompt and the statistically significant 2-way Age by Prompt interaction for the by-participant analyses cannot be generalised.

Lastly, a Spearman's correlation showed that ranking and rating data were highly statistically significantly correlated ($\rho = -0.976$, $p<0.0005$).

8.3.3 DISCUSSION OF EXPERIMENT 6 RESULTS

The Effect of Age on the Subjective Likeness of Composite Images Produced (aim (v))

Overall, results supported the hypothesis that there would be a general improvement in the quality of the *likeness* of the composites produced with age. Composites produced by adults were evaluated as statistically significantly better likenesses (as measured by both ranking and rating tasks) than composites created by children of all ages. Additionally the composites produced by 10-year-olds were evaluated as statistically significantly better likenesses than composites created by 6-year-olds. However, there

was no statistically significant difference between likeness scores for composites produced by 6- and 8-year-old or 8- and 10-year-old children.

Of particular interest is the performance of the youngest participants (aged 6-years) as this is the age group that has received little attention to date. The findings of the present study are in contrast to the research findings of Davies *et al.* (1989) who using a naturalistic event, showed that there was no statistically significant effect of age on the likeness rating of composites created by children aged 6- to 7-years and 10- to 11-years. However, the present study included a younger group of 6-year olds (aged 5- to 6-years rather than aged 6- to 7-years). The findings of the present study are also in contrast to the findings of Schwartz-Kenney, Norton, *et al.* (1996) who included this younger age group of 5- to 6-year-olds in their naturalistic study and found a statistically significant difference between this younger age group of children and 8- to 9-year-olds using the Identikit system. When considering the actual likeness scores composites created by children received comparisons can be made with the research studies described in Appendix Va. In the present study composites constructed by 6-year-olds were provided with a mean likeness rating of almost 30%. This was approximately the same as the 31% likeness rating composites constructed by 6-7 year olds were awarded in the study by Davies *et al.* (1989) and lower than the 35% rating composites constructed by 5- to 6-year-olds were awarded in the study by Schwartz-Kenney, Norton *et al.* The 40% likeness rating awarded to composites constructed by 8 year olds in the present study was also slightly lower than the 42% composites constructed by 8-9 year olds in Schwartz-Kenney, Norton *et al.*'s study. Finally, the 45% likeness rating awarded to composites constructed by 10-year olds in this study matched the 45% awarded to composites constructed by 10 year olds in the Davies *et al.* study.

As mentioned in Section 8.1, differences between studies (including the present study) may be due to the different circumstances composites were evaluated under, as will be considered now. Possible reasons for the lower likeness scores in the present study compared to Schwartz-Kenney, Norton *et al.* may simply be explained by the fact that likeness ratings involve subjective scores and in the present study children's composites were included together with composites constructed by adults. Participants providing the ratings were therefore comparing children's and adults composites on the same scale, whereas no adult comparison group was included in Schwartz-Kenney, Norton *et al.*'s study. Furthermore, both Davies *et al.* and Schwartz-Kenney, Norton *et al.* employed 5 and 6 point likeness scales (respectively) whereas the present study utilised a 10-point scale, which would provide a larger range of likeness scores.

Although composites created by the younger age groups of children received lower likeness ratings than one of the previous studies described above, a recent study conducted by Memon using the E-FIT system provides a different pattern of results, which are lower than in the present study (A.Memon, personal communication, 14th May 2004). Children aged 7- and 11-years took part in a ten-minute interaction with a stranger. Following a one-day delay children were interviewed using the CI and then constructed a facial composite using the E-FIT system with an interviewer who had received training via a specialist E-FIT operator. Complete interviews lasted for approximately 20 minutes. When rated for likeness on a 10-point scale, results showed composites constructed by 7-year-olds were awarded 28% and composites constructed by 11-year-olds were awarded 35%, and this difference was not statistically significant. Ratings for both of these groups were lower than those reported in the present study and in the previous studies cited. A direct comparison cannot be made between the present study and Memon's study due to methodological differences. However, it is interesting

to note that using the prompts as an appropriate interview technique in the present study resulted in composites which received higher ratings on average than those constructed following a CI.

Throughout the results, variations within the age groups were illustrated. Therefore, although the data showed a consistent increase in the *ranking* of composites with age, composites produced regardless of age received the whole range of ranking scores (from best to worst) and over 25% of composites created by 6-year olds were ranked within the top two groups; this figure was almost 50% for composites created by 8-year-olds and over 50% for composites created by 10-year-olds. Again, although the data showed a consistent increase in the *rating* of composites with age, composites produced regardless of age received almost the whole range of rating scores (from best to worst). Around 15% of composites constructed by 6 year olds received a rating of 6 and over, this figure was over 25% for composites constructed by 8-year-olds and over 35% for composites constructed by 10-year-olds.

In response to the questionnaire survey in Study 1, the majority of operators stated that the youngest age of child they would be willing to interview was 10-years of age. The results from the likeness evaluations of the present study indicate that if operators are willing to interview 10-year-old child witnesses then there is no reason to automatically discount children as young as 5- to 6-years from constructing composites.

Differences between Using the Visual or the Verbal Prompts on the Subjective Likeness of Composite Images Produced (aim (vi))

The results of Experiment 5 showed that descriptions were elicited significantly faster using the visual prompts. However, although composites constructed by participants using the visual prompts were evaluated as better quality (in terms of ranking and rating

tasks) than composites created using the verbal prompts, the difference was not statistically significant.

The positive effects of the visual prompts on the quality of children's descriptions (as measured by accuracy and inaccuracy) were not extended to the composite construction process possibly due to the small effect sizes reported for Prompt in Study 3 (only 6.8% of the overall variation in accuracy, and only 6.1% of the overall variation in inaccuracy, was attributable to the influence of Prompt). This again highlights the large variations within the data.

Finally, there are a number of reasons, including some related to the methodological design of the experiment, which may have contributed to the findings related to age and prompts described. These are discussed in detail in Section 8.5.2.

8.4 EXPERIMENT 7 - FORCED CHOICE MATCHING TASK

8.4.1 METHOD

Design

This phase employed a within-participants design, containing two factors. The factor 'Prompt' had two levels (Visual and Verbal) and the factor 'Age' had four levels (6-, 8- and 10-years and Adult). The dependent variable was the number of correct identifications in a matching task, where the participants were required to identify the corresponding target from a photo-spread whilst viewing a composite.

A number of randomisations were employed to control for order effects. The order in which the composites were presented was randomised for every participant. The number of composites from each condition was counterbalanced across participants, such that

each condition and target appeared equally frequently. Two line-ups were created for the matching task where the positions of the targets were altered between line-ups. The presentation of the two line-ups was counterbalanced.

Participants

Twenty adults (six males and 14 females) were drawn from students, staff and visitors to The Open University. They were recruited informally and were not paid for their participation. Their ages ranged from 24 years to 60 years. All participants spoke English as their first language. None of the participants had taken part in the construction of the composites (Experiment 5) or the ranking and rating of the composites (Experiment 6).

Materials

A ten-person simultaneous line-up was constructed from the five original target photographs (taken from the identification line-up task in Experiment 5) and five similar distractor photographs in the same format (taken from the distractors in the line-ups used in Experiment 5)⁸⁹. Portrait colour prints (7cm by 5.5cm) were placed on an A3 sheet of paper lettered A to J in 2 rows. Two line-ups were created using the same photographs with the order of the targets and distractors altered. Each line-up can be seen in Appendix Vj.

The 80 final composites, constructed for Experiment 5, were printed in greyscale at 72 d.p.i. resolution (E-FIT's maximum output), and measured approximately 15cm by 10cm (Appendix Ve shows each composite constructed). Each composite was numbered from 1 to 80. Participants were given an instruction sheet to explain the task

⁸⁹ The line-up consisted of two 'type likenesses' which shared common characteristics (based on Face Shape, Hair Colour, Hair Type, Hair Length and Hair Style).

required of them (Appendix Vk)⁹⁰. Finally, participants were provided with a response sheet (Appendix VI) to indicate their selections from the line-up. This contained a column containing composite numbers and columns to indicate which photograph they chose and how confident they were in their matching decision.

Participants were also given a response sheet in the form of a table with a column for the photograph selected (A-J). The general procedure followed the instruction sheet and is summarised below:

Procedure

The experiment was conducted over two days at least one week apart. Participants were instructed that they were about to see a set of face images which would consist of a photographic line-up of ten men (labelled A to J) and 20 E-FIT (facial composite images) (numbered between 1 and 80). Participants were instructed that for each composite they should indicate on the response sheet the number of the composite and the letter of the man from the photographic line-up they thought it most resembled.

It was stressed to participants that the numbers (1 to 80) and letters (A to J) were simply identifying letters to allow them to record responses and that the letters gave no indication as to the correct response. Participants were also informed that there was no time limit; that each judgement should be independent of any others; that they could choose a photograph as many or as few times as they wished; that they did not have to choose every photograph; and that the target from which each composite was constructed would always be present in the array.

In the second phase of the experiment, participants were presented with the alternate photographic line-up and a further 20 composites (again consisting of an equal number

⁹⁰ The instruction sheet included some brief background information as detailed in Appendix Vg.

form each condition). They then completed the task described above again. At the end of the task, participants were thanked and de-briefed.

8.4.2 RESULTS

This section considers data for the matching task. As with the previous results section data is analysed first by-item (by composite) and then by-participant.

Effects of Target

The effects of Target on the matching of the E-FITs were tested for. A 5x4x2 between participants ANOVA design was employed. The factor Target had five levels (Targets 1 to 5), the factor Prompt had two levels (Visual and Verbal), and the factor Age had four levels (6-, 8- and 10-years and Adult). Results showed the main effect of Target, the 2-way Target by Age interaction and the 3-way Target by Age by Prompt interaction were all statistically non-significant. However there was a statistically significant 2-way Target by Prompt interaction. This was due to the total number of correct matches being similar for both the verbal and visual prompts for Targets 1 and 3. However, for Targets 2 and 4 the total number of correct matches for the verbal prompts was higher than for the visual prompts and the opposite was true for Target 5. This point will be considered in the discussion and Target will not be considered further in the results section. Full ANOVA calculations are provided in Appendix Vi.

Matching Data (considered by-item)

It should be noted that each composite was matched to the ten image array by ten participants. Therefore, by chance alone a composite should be correctly matched once. The mean number of participants correctly matching each composite correctly to its target across all participants and all conditions (Age and Prompt) was 1.99. 53 out of 80 composites were correctly matched-to-target by at least one participant (66.25%).

The highest number of correct matches for a composite was 10 (maximum number of correct matches possible = 10). This was for a composite created by an 8-year-old using verbal prompts (i.e. 100% of the participants who viewed the composite correctly matched it to its target photograph). Figure 8.5 shows this E-FIT along with its corresponding target.

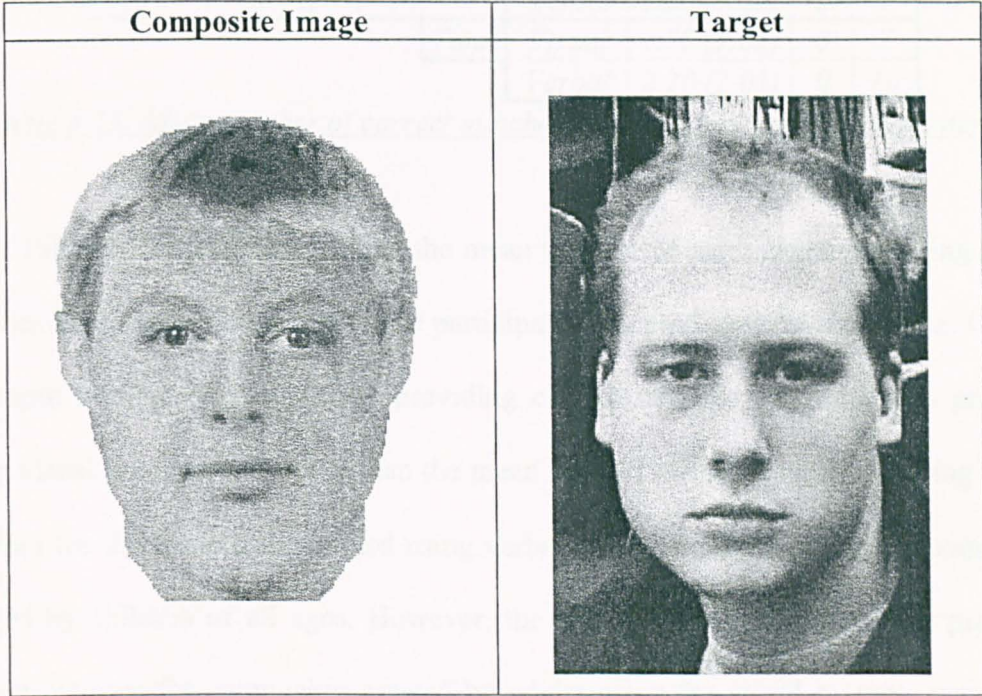


Figure 8.5: E-FIT of Target 2 correctly matched by all participants in Experiment 7

Table 8.13 summarises the mean number of correct matches when considered by-item (collapsed across participant). It should be noted that the maximum score both by Age and by Prompt is 10, and the number of correct matches which would be made by chance is 1.

Age	Mean (<i>S.D</i>)	Min	Max	Prompt	Mean (<i>S.D</i>)	Min	Max
6 years	0.80 (1.01)	0	3	Visual	0.60 (0.70)	0	2
				Verbal	1.00 (1.25)	0	3
8 years	1.45 (2.61)	0	10	Visual	0.90 (1.52)	0	4
				Verbal	2.00 (3.37)	0	10
10 years	1.85 (1.87)	0	6	Visual	1.50 (1.84)	0	6
				Verbal	2.20 (1.93)	0	6
Adult	3.85 (1.81)	1	7	Visual	4.10 (1.79)	2	7
				Verbal	3.60 (1.90)	1	7
			Total	Visual	1.78 (2.03)	0	7
				Verbal	2.20 (2.03)	0	10

Table 8.13: Mean number of correct matches (by-item analyses) in Experiment 7

From Table 8.13 it can be seen that the mean number of participants providing correct matches increased with the age of the participant who produced the composite. Overall, the mean number of participants providing correct matches to composites produced using visual prompts was lower than the mean number of participants providing correct matches for the composites created using verbal prompts. This was true for composites created by children of all ages. However, the mean number of participants providing correct matches for composites created by adults using the visual prompts was higher than for composites produced by adults using the verbal prompts. Although composites created by 6-year-olds were being matched, on average, below chance level, the actual percentage of composites constructed by 6-year-olds matched at below chance level was 50%. Thirty per cent of composites constructed by 6-year-olds were matched at chance level and 20% were matched above chance level. Additionally, although composites created by 8-year-olds using the visual prompts were being correctly matched, on average below chance level, the actual percentage of composites constructed by 8-year-olds using the visual prompts matched at above chance level was 30%.

It was hypothesised that there would be a general improvement in the utility of composites produced with age. A 4x2x2 between participants ANOVA design was

employed. The factor Age had 4 levels (6-, 8- and 10-years and Adult), the factor Prompt had 2 levels (Visual and Verbal), and the factor Identification had 2 levels (Correct and Incorrect)⁹¹. The dependent variable was the total number of correct matches for each composite. Results showed the main effect of Prompt was statistically non-significant [$F_{(1,64)}=0.107$, $p=0.745$] and the 2-way Age by Prompt interaction was also statistically non-significant [$F_{(3,64)}=0.098$, $p=0.961$]. There was a statistically significant main effect of Age [$F_{(3,64)}=9.574$, $p<0.0005$, partial $\eta^2=0.310$] suggesting that older participants produced composites that were correctly matched more often and supporting the hypothesis stated above. This effect was further analysed using a post-hoc Scheffé test, the results of which are shown in Table 8.14.

Age 1	Age 2	Mean Difference	Significance
6 years	8 years	-0.650	$p=0.777$
	10 years	-1.050	$p=0.418$
	Adult	-3.050	$p<0.0005$
8 years	6 years	0.650	$p=0.777$
	10 years	-0.400	$p=0.936$
	Adult	-2.400	$p<0.005$
10 years	6 years	1.050	$p=0.418$
	8 years	0.400	$p=0.936$
	Adult	-2.000	$p<0.05$
Adult	6 years	3.050	$p<0.0005$
	8 years	2.400	$p<0.005$
	10 years	2.000	$p<0.05$

Table 8.14: Post hoc Scheffé test of the effect of Age on the matching scores in Experiment 7

The results shown in Table 8.14, together with the data described above, suggest that there were statistically significantly more correct matches for composites constructed by adult participants than for composites constructed by children. Interestingly, the difference between the mean numbers of correct matches for composites produced by 6-, 8- and 10-year-olds was statistically non-significant.

⁹¹ The between participants factor Identification will be considered at the end of Section 8.4.2.

Matching Data (considered by-participant)

It should be noted that each participant viewed 40 composites. Therefore, by chance alone a participant should correctly match 4 composites. On average, participants scored well above chance and the mean number of composites matched correctly to target was 8 (i.e. on average, 20% of the composites viewed by a participant were correctly matched to its target photograph). The highest number of correct matches for a participant was 12 (30%).

Table 8.15 summarises the mean number of correct matches when considered by-participant (collapsed across item). It should be noted that the maximum score by Age is 10 and chance is 1. The maximum score by Prompt is 5 and chance is 0.5.

Age	Mean (<i>S.D</i>)	Min	Max	Prompt	Mean (<i>S.D</i>)	Min	Max	
6 years	0.8 (1.26)	0	2	Visual	0.30 (0.57)	0	2	
				Verbal	0.50 (0.69)	0	2	
8 years	1.5 (1.69)	0	4	Visual	0.45 (0.69)	0	2	
				Verbal	1.05 (1.00)	0	4	
10 years	1.85 (1.51)	0	3	Visual	0.75 (0.72)	0	2	
				Verbal	1.10 (0.79)	0	3	
Adult	3.85 (1.94)	0	4	Visual	2.05 (0.76)	0	3	
				Verbal	1.80 (1.15)	0	4	
				Total	Visual	0.89 (0.97)	0	3
					Verbal	1.11 (1.01)	0	4

*Table 8.15: Mean number of correct matches (by-participant analyses)
in Experiment 7*

From Table 8.15 it can be seen that the pattern of results is the same as the by-item data (Table 8.13). A 4x2 within-participants ANOVA design was employed. The factor Age had 4 levels (6-, 8- and 10-years and Adult), and the factor Prompt had 2 levels (Visual and Verbal). The dependent variable was the total number of correctly matched composites. The ANOVA revealed the same pattern of results as that conducted above by-item i.e. a statistically significant main effect of Age [$F_{(3,57)}=40.400$, $p<0.0005$, partial $\eta^2=0.680$] (again supporting the hypothesis that there would be a general

improvement in the utility of composites produced with age), a statistically non-significant main effect of Prompt [$F_{(1,19)}=1.605$, $p=0.221$] and a statistically non-significant 2-way interaction between Age and Prompt [$F_{(3,57)}=1.904$, $p=0.139$].

Using the results from the by-item and by-participant analyses min F' was calculated for the effects. The results of the Min F' and degrees of freedom calculations show a similar pattern of results as the previous calculations for the by-item and by-participant analysis i.e. a statistically significant main effect of Age [$F_{(3,103.739)}=7.722$, $p<0.05$] (supporting the hypothesis that there would be a general improvement in the utility of composites produced with age), a statistically non-significant main effect of Prompt [$F_{(1,101.514)}=0.5088$, $p>0.05$] and a statistically non-significant Age by Prompt interaction [$F_{(3,99.789)}=0.639$, $p>0.05$]. Therefore, the statistically significant main effect of Age can be assumed to generalise to the whole population and to a different set of items (composites) (i.e. it is not that there is a particular composite or participant unduly affecting the results).

Incorrect Matches

The number of selections of each photograph in the 10 person line-up (correct match of the target and incorrect selection of a foil) was calculated as a percentage of the total number of selections for each target. These data are shown in Table 8.16. Correct selections are shown in bold italics.

Photo	Target				
	1	2	3	4	5
1	19.38%	3.75%	14.38%	13.75%	4.38%
2	5.63%	23.13%	7.50%	4.38%	13.13%
3	6.88%	10.00%	13.75%	16.88%	6.88%
4	24.38%	5.00%	20.00%	27.50%	2.50%
5	1.88%	8.13%	1.25%	0.63%	18.13%
6	10.63%	3.13%	7.50%	5.63%	2.50%
7	9.38%	10.63%	8.13%	10.00%	14.38%
8	14.38%	15.00%	8.13%	12.50%	13.13%
9	3.75%	15.63%	11.25%	5.63%	12.50%
10	3.75%	5.63%	8.13%	3.13%	12.50%

Table 8.16: Percentage of selections of each array photograph in Experiment 7

Table 8.16 shows that for Targets 2, 4 and 5 the most selected photograph was the corresponding target (i.e. a correct match). For Target 1 photograph 4 (a foil) was the most frequently selected and the correct target was matched second. For Target 3, again photograph 4 (a foil) was the most frequently selected, followed by photograph 1 (a foil) with the correct target matched third.

A possible explanation for these results is that targets 1, 3 and 4 could be described as a “type likeness” of each other. The photographs of targets 1, 3 and 4 in Appendix Vf show that all of these targets share the common characteristics of short dark hair, dark eyebrows and long, narrow faces and noses. Figures 8.6 and 8.7 illustrate this explanation by providing examples of E-FITs constructed of targets 1 and 3 respectively which were incorrectly matched to target 4.




Composite Image	Most frequently chosen foil	Correct target
	 Target 4	 Target 1

Figure 8.6: E-FIT of Target 1 frequently incorrectly matched to Target 4 in Experiment 7




Composite Image	Most frequently chosen foil	Correct target
	 Target 4	 Target 3

Figure 8.7: E-FIT of Target 3 frequently incorrectly matched to Target 4 in Experiment 7

In Figure 8.6, the composite image was created by a 10-year-old child using the verbal prompts. Nine of the ten participants in the forced choice matching task matched this composite incorrectly to Target 4. In Figure 8.7, the composite image was created by an adult using the verbal prompts. Six of the ten participants in the forced choice matching task matched this composite incorrectly to Target 4.

Identification

Finally, participants' identification performance on the nine-person target present identification task following their composite construction was considered in more detail.

Table 8.17 shows the mean rank and mean rating scores for composites produced by participants who made either a correct or an incorrect identification in Experiment 5.

Identification	Mean		
	Rank (<i>S.D.</i>)	Rating (<i>S.D.</i>)	N°. correct matches (<i>S.D.</i>)
Correct	2.40 (0.87)	4.65 (1.99)	2.25 (2.31)
Incorrect	2.68 (0.84)	4.02 (1.76)	1.52 (1.94)

Table 8.17: Mean rank and rating scores for correct and incorrect identifications for Study 4

From Table 8.17 it can be seen that the mean rank score for composites produced by participants who made correct identifications was slightly lower (therefore, indicating a better rank) than the mean rank score for composites produced by participants who made incorrect identifications. The 4x2x2 between participants ANOVA (employed on the factors specified in the by-item ranking results section) conducted on the ranking data showed a statistically non-significant main effect of Identification [$F_{(1,64)}=0.210$, $p=0.648$] and the 2-way interactions between Identification and Age [$F_{(3,64)}=0.106$, $p=0.956$]; Identification and Prompt [$F_{(1,64)}=0.433$, $p=0.513$]; or 3-way Age by Prompt by Identification [$F_{(3,64)}=0.175$, $p=0.913$] were statistically non-significant.

From Table 8.17 it can be seen that the mean rating score for composites produced by participants who made correct identifications was slightly higher the mean rating score for composites produced by participants who made incorrect identifications. The 4x2x2 between participants ANOVA (employed on the factors specified in the by-item rating results section) conducted on the rating data showed the main effect of Identification

was statistically non-significant [$F_{(1,64)}=0.262$, $p=0.610$]. The 2-way interactions between Identification and Age [$F_{(3,64)}=0.198$, $p=0.898$]; Identification and Prompt [$F_{(1,64)}=0.066$, $p=0.798$] or 3-way Age by Prompt by Identification [$F_{(3,64)}=0.075$, $p=0.973$] were statistically non-significant.

From Table 8.17 it can also be seen that the mean number of correct matches for composites produced by participants who made correct identifications was higher than the mean number of correct matched for composites produced by participants who made incorrect identifications. The 4x2x2 between participants ANOVA (employed on the factors specified in the by-item matching results section) conducted on the matching data showed the main effect of Identification was statistically non-significant [$F_{(1,64)}=0.886$; $p=0.350$]. The 2-way interactions between Identification and Age [$F_{(3,64)}=1.137$; $p=0.341$] Identification and Prompt [$F_{(1,64)}=0.242$; $p=0.624$] and 3-way Age by Prompt by Identification interaction [$F_{(3,64)}=0.603$; $p=0.615$] were statistically non-significant.

Lastly, the data from Experiment 7 were entered into Spearman's correlations with the data from Experiment 6 (ranking and rating tasks). The average ranking and the number of participants correctly matching a composite to its corresponding target were statistically significantly correlated [$\rho=-0.655$, $p<0.0005$] and the average rating was also statistically significantly correlated with the number of participants correctly matching a composite to its corresponding target [$\rho=0.644$, $p<0.0005$].

8.4.3 DISCUSSION OF EXPERIMENT 7 RESULTS

The Effect of Age on the Objective Utility of Composite Images Produced (aim (vii))

Overall, the results supported the hypothesis that there would be a general improvement in the *utility* of the composites produced with age. For this objective measure of matching, again the composites produced by adults were statistically significantly different from the composites created by all ages of children. However, there was no statistically significant difference between the correct matching for composites produced by 6-, 8- or 10-year-old children. This finding is consistent with research by Flin *et al.* (1989) who found a statistically significant difference between adults and children's Photo-FITs but did not find a statistically significant difference between 8- to 9-year-olds and 11- to 12-year-olds (although this difference approached significance).

The by-item analyses showed that in total, over 66% of composites were correctly matched to target by at least one participant and the only composite to get matched by all participants was produced by an 8-year-old. On average, the number of participants matching composites correctly to target was above chance⁹². Approximately 20% of participants matched each composite correctly and this figure ranged from 8% to 14.5% to 18.5% to 39% of participants correctly matching composites constructed by 6-, 8-, 10-year-olds and adults respectively. The figure for the adult participants is consistent with research conducted using the E-FIT system by Turner (2004), who using a similar matching task reported between 29% and 43% of participants correctly matching composites created by adults.

⁹² The by-participant analyses also showed that on average, participants scored above chance and approximately 20% of the composites viewed by a participant were matched to the correct target photograph.

Only one study has been published which examined the utility of children's composites. As described in Appendix Va, Flin *et al.* (1989) asked 20 judges to sort composites constructed by children and adults into six piles corresponding to likenesses of six target faces. Results showed that composites were correctly sorted on average by 34% of participants for composites created by 8- to 9-year-olds, 40% of participants for composites created by 11- to 12-year-olds and 50% of participants for composites created by adults. The percentage of correct matches reported by Flin *et al.* are higher than the averages in the present study. Additionally, Flin *et al.* found that on average, composites created by participants of all ages were correctly sorted above chance. In contrast, the present study found that although on average the number of participants matching composites correctly to target was above chance, when comparing age and prompt conditions, composites created by 6-year-olds, and composites constructed by 8-year olds using the visual prompts were being matched, on average, below chance level.

The contrasting findings of the present study with those of Flin *et al.* may first be explained by methodological differences. The younger age group of 6-year-olds were not included in Flin *et al.*'s study. Additionally, Flin *et al.* did not include a delay in their study between observation of the target and composite construction. Furthermore, participants in Flin *et al.*'s study had the option to add artistic enhancement to refine their composites. Research has shown that artist elaboration can have an impact on overall image quality (Gibling and Bennett, 1994) and this option was not provided in the present study.

Secondly, different evaluation tasks were used in each study. Flin *et al.* asked judges to sort 144 composites between the six target faces used in their experiment. The present study employed the more difficult task, of asking judges to match 40 composites to a

ten person line-up which included the five targets used in the study along with five distractor faces. This was more similar to the task employed by Turner (2004). As described in Section 8.4.1, the line-up consisted of two 'type likenesses' and as a result many of the targets were confused. This point is illustrated by the number of consistently 'incorrectly' selected targets. As described in Chapter 4, a facial composite is intended to be a 'type likeness' rather than a photographic representation. A possible interpretation is that a number of the E-FIT composites that were incorrectly matched in the present study may have been good 'type likenesses' which would have prompted a number of possible matches in an actual police appeal.

Throughout the results of the present study variations within the age groups were illustrated. Therefore, although composites created by 6-year-olds were being matched, on average, below chance level, 30% of composites constructed by 6-year-olds were matched at chance level and 20% were matched above chance level. As concluded from the likeness data, the matching data also indicates that if operators are willing to interview 10-year-old child witnesses (as they stated in response to the questionnaire survey) then there is no reason, on the basis of these results, to automatically discount children as young as 6-years from constructing composites.

Differences Between Using the Visual or the Verbal Prompts on the Objective Utility of Composite Images Produced (aim (viii))

The results of Experiment 5 showed that descriptions were elicited significantly faster using the visual prompts. However, although composites constructed by participants using the visual prompts were evaluated as better quality (in terms of the matching tasks) than composites created using the verbal prompts, the difference was not statistically significant. As described previously in Section 8.3.3, the positive effects of the visual prompts on the quality of children's descriptions were not extended to the

composite construction process possibly due to the small effect sizes reported for Prompt in Study 3, again highlighting the large variation within the data.

The statistically significant 2-way Prompt by Target interaction described suggests that perhaps different faces could more easily be described using the visual prompts and some using the verbal prompts (in terms of the number of correct matches for the subsequent composite constructed). This perhaps reflects a number of adult participants comments made during Study 3: a proportion of adults stated that they preferred visual prompts for some features and verbal prompts for other features.

Finally, the ranking, rating and matching score for composites constructed by participants who went on to make a correct identification were higher, on average, than for those who went on to make an incorrect identification. However, the differences were not statistically significant. The identification task was included in the present study to determine whether participants had attended to the target and as described in Section 8.2.3 there appeared to be a group of children who had simply failed to attend to the target (55% of 6-year-olds made an incorrect identification). It is possible that these children were simply not encoding the target face and subsequently could not recognise it. Children may have simply encoded the face at a level which enabled them to produce a 'type likeness' which was enough to construct a composite but not enough to correctly identify the target in an identification task. This would also explain the low numbers of correct matches in the matching evaluation task for composites constructed by children compared to adults (where participants were asked to match composites to a photograph within an array of type likenesses). Alternatively, the identification task may not be telling us whether children encoded the face or not as there was no statistically significant difference between the quality scores awarded to composites (as measured

by ranking, rating and matching tasks) and participants' subsequent identification performance.

As described previously, there are a number of reasons, including some related to the methodological design of the experiment, which may have contributed to the findings related to age and prompts, which are discussed in the following section (Section 8.5.2).

8.5 GENERAL DISCUSSION OF STUDY 4

The main aim of Study 4 was to compare the use of a set of visual prompts and verbal prompts as an appropriate interview technique for adult and child participants' composite constructions of unfamiliar faces. Three experiments were conducted. Analyses of results initially considered whether adults and children could construct composites using the prompts in an experimental setting. Analyses also considered whether a number of findings from Experiment 4 would be replicated, including: the time taken to obtain prompted descriptions; participants' use of "don't know" prompts; and participants' performance at an identification task following composite construction. Finally, analyses included an evaluation of both the subjective likeness and the objective utility of the composites produced by both adult and child participants.

8.5.1 SUMMARY OF STUDY 4 RESULTS

The most important observation to make is that the collective findings from Study 4 illustrate that children from the age of 6-years can produce facial composites of an unfamiliar face using a set of visual or verbal prompts and the computerised composite system E-FIT. However, when composites were considered in terms of subjective and

objective evaluation measures it became clear that the task drew on developmentally sensitive skills. Overall results supported the hypothesis that there would be a general improvement in the quality of the *likeness* and the *utility* of the composites produced with age. Although, as discussed, there were large variations within the data throughout all of the results sections, and differences within age groups were of equal or greater importance than difference between age groups. That is, composites produced by adults and older children were not always evaluated as better and younger children's composites were included amongst the best likenesses and utility rated by independent judges.

Of particular interest was the performance of the youngest participants (aged 6-years) as this is the age group that has received little attention to date. Results showed that composites produced regardless of age received a wide range of ranking and rating scores (from best to worst) and over 25% of composites created by 6-year olds were ranked within the top two groups. Around 15% of composites constructed by 6 year olds received a rating of 6 and over.

There was no statistically significant difference between the *utility* scores for composites produced by 6-, 8- or 10-year-old children. On average, the number of participants matching composites correctly to target was above chance. However, the composites created by 6-year-olds were being matched, on average, below chance level. Therefore it is again important to highlight the large variation *within* the age groups. For example, although composites created by 6-year-olds were being matched, on average, below chance level, 50% of composites constructed by 6-year-olds were matched at chance level or above. The utility scores of the present study were lower than the findings of the only other study which has examined the objective utility of composites

constructed by E-FIT. However, a direct comparison between the two studies cannot be made due to the methodological differences in composite construction and evaluation described in Section 8.4.3.

As argued in the previous discussion sections of this chapter (Sections 8.2.3 and 8.3.3) the results from the likeness and utility ratings of the present study indicate that if operators are willing to interview 10-year-old child witnesses then there is no reason to automatically discount children as young as 6-years from constructing composites.

Results also suggested that if participants had been provided with unlimited time to construct a composite, adults and older children may have spent longer and made further changes. Whereas, the composites obtained from the younger age groups were possibly as complete as they would ever be, and that allowing more time may not have resulted in significantly different composites. It was argued that these findings provide support for the quantitative change in the development facial processing described in Section 3.2.2 i.e. younger children did not want to make as many changes to the composite they were presented with. The holistic system E-FIT also allows witnesses to make configural changes to a composite, and perhaps younger children could not go on to make such changes.

As described previously, the positive effects of the visual prompts on the quality of children's descriptions were not extended to the composite construction process. Although the results of the present study showed composites constructed by participants using the visual prompts were evaluated as better quality (in terms of likeness and utility) than composites created using the verbal prompts this difference was not statistically significant. This was possibly due to the small effect sizes reported for

Prompt in Study 3, which emphasises the large variation within the data. With a larger number of participants, or perhaps under different circumstances some significant differences between the composites constructed using the different sets of prompts may possibly be observed.

An identification task was included in the present studies to determine whether children had attended to the target. Results showed that all participants performed at above chance level in the identification task following composite construction. As described it is possible that some children were only encoding the target face at a ‘type likeness’ level which enabled them to construct a composite but not enough to correctly identify the target in an identification task.

Finally, the findings replicated some of the positive findings from Experiment 4: descriptions were elicited significantly faster using the visual prompts for child participants; and children were comfortable selecting both visual ‘?’ and verbal “don’t know”.

As noted previously, there are a number of reasons, including some related to the methodological design of the experiment, which may have contributed to the findings which are now discussed.

8.5.2 LIMITATIONS OF STUDY 4

There are a number of methodological limitations which may have contributed to the findings described.

The imposed time limit of thirty minutes to interview and construct a composite image may have masked or exacerbated differences between composites created by

participants of different ages and/or between composites created using the visual or verbal prompts. Thirty minutes is a very short time to obtain a description and create an accurate likeness. As described, the data from the Questionnaire Survey (Chapter 5) illustrated that average time taken to conduct an interview with a child witness (from description to composite construction) was just over two hours, with a range from an hour and five minutes to three and a half hours.

System limitations may also have had an effect on the results. Although the computerised E-FIT system is more flexible, provides a wider range of features, and allows a better blending of features than the earlier non-computerised composite systems, the operator did not use a specialised graphics package with which to refine their composites and artist elaboration can have an impact on overall image quality (Gibling and Bennett, 1994).

All composite systems and other recall techniques require an operator or technician to produce an image and the quality of the composite may be impaired by factors associated with the operator as well as the witness. In the present study the operator received training via the E-FIT manual and observations of experience E-FIT operators. However, there is evidence that the quality of a composite can be influenced by the skill and the experience of the operator (e.g. Davies *et al.*, 1983). Koehn and Fisher (1997) showed that likenesses made by an experienced operator with the aid of reference photographs were recognised 77% of the time. Memon (2004) also found a significant effect of operator on the likeness rating of composites constructed using the E-FIT system (A.Memon, personal communication, 14th May 2004). The operator in the present study only had limited training in the use of the E-FIT system. Therefore

potential practice effects were controlled for by counterbalancing the order of the age of participant interviewed.

No typical interview structure could be included as a control condition in Experiment 5 for child participants due to the lack of interview structure employed by operators (as determined by the responses to the questionnaire survey in Study 1).

As described in Section 8.4.1 the target faces used in the matching task line-up consisted of two-type likeness. This was due to the two type likenesses of the five target faces using the present study. As the photographs of the targets used in Appendix Vf show, Targets 1, 3 and 4 were one type likeness and Targets 2 and 5 were another type likeness, in terms of having similar Face Shapes, Hair Colour, Hair Type, Hair Length and Hair Style. The use of only two type likenesses rather than five very different target faces may have affected the number of correct matches made during the objective evaluation of the composites as described previously.

Finally, as with Study 3, due to restrictions from the school used in the present study audio (rather than video) recording was used, supplemented by notes made by the author throughout each interview. If video recording were possible it would be interesting to investigate a number of measures which were briefly noted by the operator throughout the interview. For example, whether participants changed any of their initial descriptions or the number of exemplars of a feature a participant looked at. A video recording would provide information which is not recorded in the E-FIT. For example, whether participants did not change a feature because the original feature selected on the basis of their description was correct (they looked at a number of exemplars for the feature in question but did not change it) or because simply because they did not remember what the feature in question looked like. A comparison of the

description of the initial E-FIT (provided by witnesses selection of the prompts) with the description of the final E-FIT (provided by E-FIT's description boxes) could also be made if a video record was available. For example, a participant may have initially selected the prompt for "light hair" and then changed the description to "dark hair" which they then went on to lighten using E-FIT's 're-tone' button. This also applies to whether participant resized or moved a feature. ABE (Home Office, 2002) also stresses that interviews with child witnesses should be video recorded, following this guidance could make the interview process forensically more valid. This is of particular importance when using prompts with children as children may not verbalise their choice of prompt and a video would illustrate the child's selection and could not be open to scrutiny if presented to a court as evidence.

All of the methodological limitations described were a result of the experimental design. However, such a design was employed in order to allow the first large scale study of children's composite construction using a computerised system and to allow a comprehensive evaluation of the composites constructed using three different techniques of measurement. The present study is an initial step in what may potentially be a whole series with children.

In summary, a number of issues have been raised which have implications for future research (some of which have been touched upon in the above section). The implications of Study 4 and future research will be considered in detail in the main Discussion (Chapter 9) and therefore will not be considered here.

CHAPTER 9: GENERAL DISCUSSION

This chapter begins with an overview of the experimental chapters described in detail in Chapters 5 to 8. A summary of the findings will be provided in line with the three research questions of the thesis and the theoretical contributions of the research. A critique of the methodology employed in the thesis is then provided. Consideration is given to future directions for research. Finally, the methodological and practical implications of the research are described before concluding comments are made.

9.1 OVERVIEW OF EXPERIMENTAL CHAPTERS

The purpose of the research presented was to gain an understanding of children's verbal descriptions of unfamiliar faces, and to explore how children might be assisted through appropriate interview techniques to provide computerised facial composite constructions of unfamiliar faces. The research comprised seven experiments grouped into four studies which can be summarised as follows:

- Study 1 involved a questionnaire survey of facial composite operators to identify their current practice with, and experiences and opinions of, child witnesses.
- Study 2 explored, for the first time in the existing literature, the content and quantity of children's verbal descriptions of unfamiliar faces in order to establish the language and terms children use. Study 2 comprised two experiments: Experiment 1 examined adult participants' selection of a set of facial composite images to create a smaller set for use with child participants in Experiment 2. Experiment 2 involved the collection and investigation of children's descriptions of unfamiliar facial composite images. The children's descriptions obtained in this study were used to produce an original

set of age appropriate verbal and visual prompts which were then used to interview children in Studies 3 and 4.

- Study 3 investigated the effect of the verbal and visual prompts as potentially appropriate interview techniques on the production of children's descriptions of unfamiliar faces. It comprised two experiments: Experiment 3 involved adult participants' selection of the prompts for target faces which remained in view. Their selections provided a measure of, and informed, the 'correct' prompts for later coding of child participants' responses in Experiment 4, which investigated child participants' selection of the prompts for target faces from memory.
- Study 4 then explored the use of the verbal and visual prompts as potentially appropriate interview techniques to assist adult and child participants' composite constructions of unfamiliar faces with the E-FIT system. Study 4 comprised three experiments: Experiment 5 involved child and adult participants' selection of prompts and construction of facial composites. Experiment 6 consisted of the subjective evaluation of the composites (using ranking and rating tasks) and Experiment 7, the objective evaluation of the composites (using a forced choice matching task).

Detailed methodological information and the principal results of each experiment are presented in Appendix I.

9.2 SUMMARY OF RESEARCH FINDINGS AND THEORETICAL CONTRIBUTIONS

Although each study included its own specific research aims, the research was designed to address the three general questions listed in Section 4.5 at the end of Chapter 4.

These questions were: (1) *How effectively can children produce verbal descriptions of an unfamiliar face?* (2) *Is there an appropriate interview technique which can be used to enhance children's facial descriptions for subsequent computerised facial composite constructions?* (3) *Can children produce computerised facial composites of an unfamiliar face when using an appropriate technique to enhance their verbal descriptions?* The following sections summarise findings from this research for each of these over-arching questions.

9.2.1 HOW EFFECTIVELY CAN CHILDREN PRODUCE VERBAL DESCRIPTIONS OF AN UNFAMILIAR FACE? (RESEARCH QUESTION (1))

Collectively, the findings of the present research demonstrate that children from the age of 6-years are able to effectively describe unfamiliar faces (in terms of the content and quantity of descriptions useful for subsequent composite construction). Free facial descriptions from children were limited by their language development. Children of all ages provided more facial descriptions when prompted. There were large variations within age groups and importantly all children, including the youngest age group, were able to describe many of the key facial features which is a necessary requirement for composite construction.

The findings from the questionnaire survey (Study 1) showed that almost all operators reported problems when obtaining verbal descriptions from child witnesses. Throughout the survey operators raised concerns about the time they spent interviewing children and related to this, children's concentration levels. Operators also frequently reported problems when obtaining facial descriptions from children due to their language development.

As was previously pointed out in Section 2.1.3, existing research has neglected to provide a detailed examination of the content of children's descriptions of unfamiliar faces at different ages. Prior work has been conducted to examine children's vocabulary for sexual experiences (Volbert and Van der Zanden, 1996) and their use and understanding of legal terms (Saywitz *et al.*, 1990). However, this research has not been extended to the area of facial recall and descriptions. Study 2 addressed this gap in the research and demonstrated the usefulness of exploring the detailed content of children's descriptions of unfamiliar faces.

Findings from Study 2 illustrated that the quantity of children's free descriptions of an unfamiliar face increased with age, consistent with research reviewed by Davies and Westcott (1999). As expected, verbal prompts increased the quantity of children's descriptions (e.g. Dent, 1992) and these were of particular benefit for the youngest age group of children (e.g. Pipe *et al.*, 1993).

A more detailed examination of the descriptions provided in Study 2 illustrated that children of all ages used a combination of both E-FIT (adult) terms and their own language and terms, including non-verbal descriptions, to describe unfamiliar faces. The majority of facial features children's descriptions did not match the existing E-FIT descriptions. Approximately half of the operators responding to the questionnaire survey reported using these to prompt child witnesses.

The comparison of children's descriptions to the descriptions used by E-FIT identified a number of patterns. Each pattern occurred across all age groups of children, suggesting a quantitative rather than a qualitative shift in the nature of children's descriptions with age. Such an interpretation is consistent with the research described in Section 3.2.2 which has examined facial processing in children (e.g. see a review by Gilchrist and

McKone, 2003). Indeed, the quantity of facial descriptions provided by children increased with age. However, there was no statistically significant effect of age on the quantity of children's facial descriptions due to the large variation within age groups: supporting the view described in Chapter 2 that age is only one factor in the determination of children's performance (e.g. Baker-Ward, 2002) and that individual differences in children's testimony may explain within age variation in children's interview performance (Bruck and Ceci, 1999).

In summary, Study 2 provides an important theoretical contribution to research on how children recall and describe unfamiliar faces. This is a significant finding which has implications for the ways in which children are interviewed. The detailed documentation of the specific language and terms used by children, and the comparison with existing adults' descriptions, may be used to assist interviewers, researchers and operators to develop more age appropriate interviewing strategies. Furthermore, the research findings provide some support for the more recent research which suggests a quantitative shift in children's facial processing.

On the basis of the findings from Study 1 and Study 2, and in order to address some of the concerns of operators raised in Study 1, a set of visual and verbal prompts were designed as a potentially appropriate interview technique to assist witnesses and operators in enhancing children's verbal descriptions of unfamiliar faces. In addition, the prompts were also designed to map directly onto the E-FIT composite system in order to further reduce the reliance on an operator 'translating' witnesses' descriptions to match the E-FIT index.

9.2.2 IS THERE AN APPROPRIATE INTERVIEW TECHNIQUE WHICH CAN BE USED TO ENHANCE CHILDREN'S FACIAL DESCRIPTIONS FOR SUBSEQUENT COMPUTERISED FACIAL COMPOSITE CONSTRUCTIONS? (RESEARCH QUESTION (2))

The findings from Study 3 suggest that the use of a set of visual prompts is an appropriate interview technique which can significantly enhance children's facial descriptions (in terms of reducing time taken and increasing the quantity and accuracy of children's descriptions). Although the youngest age group of children provided fewer descriptions, which were of lower accuracy, than the older children, they still provided descriptions that may have been usable as the basis of subsequent composite construction.

Operators' responses to the questionnaire survey illustrated that there was no 'typical' interview procedure being used with child witnesses, and the majority of operators reported that the interview procedure they used to obtain descriptions and composites from children was not optimal. Over half of the operators who responded to the survey described using additional materials with child witnesses in order to enhance their descriptions and subsequent facial composite constructions, and in the final part of the questionnaire approximately half of operators stated that they would like to receive a set of guidelines or prompts for use with witnesses.

Study 3 explored the use of sets of visual and verbal prompts as potentially appropriate interview techniques to assist witnesses and composite operators. As with Study 2, the results showed that children of all ages only provided limited free descriptions, especially descriptions useful for composite construction. A consistent and significant improvement in the quantity of descriptions provided was evident when using the prompts for all age groups of children, and children who did not provide any free

descriptions could provide accurate descriptions when using the prompts. Such findings are consistent with the research cited in Section 2.1.2 which proposed that even if a child has encoded an event they have difficulty spontaneously retrieving the information from memory (Salmon, 2001) and are less flexible in generating and using strategies to search their memory efficiently (Flavell, *et al.*, 1993). Relying on the reduced demands of the prompts as a recognition task reduces the demands of a recall task (Pipe *et al.*, 2002).

For children of all ages the prompted descriptions provided using the visual prompts were statistically significantly more accurate than those provided using the verbal prompts, thus illustrating that the linguistic skills needed to use the visual prompts were not as sophisticated or complex as those needed for accurate verbal recall (Gabarino and Stott, 1992; Pipe *et al.*, 2002). Furthermore, children of all ages were statistically significantly faster when selecting the visual prompts than the verbal prompts. These findings suggest that the visual prompts are an appropriate interview technique which can be used to significantly enhance children's facial descriptions, in terms of quantity and accuracy, and to reduce interviewing times, which are a key factor when working with young children.

In terms of the effect of the different age groups, the 8- and 10-year olds provided a statistically significant greater quantity of descriptions, which were more accurate than the descriptions provided by 6-year-olds. It was the 10-year-olds who benefited most from the visual prompts (as they had the largest difference between the accuracy and inaccuracy scores for their selection of the visual and verbal prompts). This is consistent with research conducted by Schwartz-Kenney, Bottoms and Goodman (1996) who reported that their line-up techniques seemed to hold most promise for use with older

children. However, it was the 6-year olds who benefited most from the visual prompts in terms of the length of time taken to obtain the prompted descriptions, possibly due to the older children's expertise at reading the verbal prompts. This consistent increase in the quantity and accuracy of descriptions provided with age again suggests a quantitative rather than a qualitative shift in children's performance with age.

As would be expected on the basis of the language research findings cited in Chapter 2 (e.g. Harley, 1996) children of all ages were statistically significantly more accurate, less inaccurate and faster when using the visual prompts compared to the verbal prompts to describe an unfamiliar target. Furthermore, the majority of children of all ages preferred responding to the visual prompts. Collectively the results are promising as they imply that the difficulties of language and concentration in obtaining children's facial descriptions described by the operators in response to the questionnaire survey can be overcome with the aid of such an interview technique.

Due to concerns which have been raised about forced choice questions (e.g. Poole and White, 1993), one of the design considerations of the prompts was that they were option-posing, rather than forced choice, by explicitly presenting children with the option of responding "don't know" (Clark *et al.*, 2001). Of particular importance was the finding from Study 3 and Study 4 that children used both the visual and verbal "don't know" prompts, and appeared to be comfortable selecting them, rather than attempting to answer questions which they did not fully understand or remember the answers to (Saywitz and Synder, 1993). The findings also provide support for the inclusion of a practice task of the use of these prompts (Schwartz-Kenney, Bottoms and Goodman, 1996).

In summary, Study 3 provides an important theoretical contribution to research on how children can be assisted through developmentally appropriate techniques to describe unfamiliar faces. Prior to this thesis there has been no explicit attempt to apply the wealth of research into the use of 'special' interview techniques to assist children's recall, such as non-verbal props and prompts (e.g. Pipe *et al.*, 1993; 2002), to the composite interview process.

9.2.3 CAN CHILDREN PRODUCE COMPUTERISED FACIAL COMPOSITES OF AN UNFAMILIAR FACE WHEN USING AN APPROPRIATE TECHNIQUE TO ENHANCE THEIR FACIAL DESCRIPTIONS? (RESEARCH QUESTION (3))

The collective findings from Study 4 illustrate that children from the age of 6-years can produce facial composites of an unfamiliar face using a set of visual or verbal prompts and the computerised composite system E-FIT, under experimental conditions. Evaluation of the composites demonstrated that the task drew on developmentally sensitive skills. However, there were large variations within age groups and composites produced by adults and older children were not always evaluated as better than younger children's. Younger children's composites were included amongst the best likenesses and were scored highly in terms of utility as rated by independent judges, indicating that children should not be excluded from producing facial composites by operators simply on the basis of their age.

All participants selected a set of visual or verbal prompts in order to provide a description of the target and produce an initial E-FIT face. When asked about the quality of this initial E-FIT, all participants asked for changes to be made, and could specify the changes to be made in order to construct a composite of the target. The complete interview (including selecting the prompts and constructing the composite) was carried out within approximately half an hour. At the end of the composite task

many participants, including all age groups of children, still wanted to make changes to the E-FIT. However, if participants had been provided with unlimited time to construct a composite it is probable that adults and older children would have spent longer and made further changes to composites. Therefore, it is unlikely that adults' and 10-year olds' optimum performance with the E-FIT system was being measured. However, the small number of younger children wanting to make further changes, and the low number of changes listed by those who did want to make changes, suggests that the composites obtained from these younger age groups were probably nearer completion and that allowing more time would not have resulted in significantly different composites. It was argued in Section 8.2.3 that these findings suggest support for the *quantitative* change in the development facial processing described in Section 3.2.2 i.e. younger children did not want to make as many changes to the composite they were presented with.

However, as described in Section 8.2.3, two factors complicate this issue and prevent any definite conclusions being drawn with regards to the relationship between the composites actually produced and those which may have been produced if unlimited time were allowed. First, it could be the case that the younger children did not fully understand the question regarding additional changes, misunderstood it or were unable to determine proactively what they might do should they be allowed more time. Second, it is far from clear that the final composite produced when unlimited time is allowed is the optimum image. There is some evidence that after a while some participant-witnesses actually make changes that introduce errors and make the composite less, not more, like the target (Pike *et al.*, 2004). Thus it cannot be assumed that the composites that the older children and adults may have gone on to produce would have actually been more accurate than those produced at the end of the time limit. Although the exact

implications of imposing a time limit are therefore impossible to draw, the data pertaining to additional changes do show that the final composite produced in this study is not indicative of the composite that would be produced if no time limits were in place.

Results showed a general improvement in the likeness and the utility of the composites produced with age. Composites produced by adults were evaluated as statistically significantly better *likenesses* (as measured by both ranking and rating tasks) than composites created by children of all ages. Additionally the composites produced by 10-year-olds were evaluated as statistically significantly better likenesses than composites created by 6-year-olds. However, there was no statistically significant difference between likeness scores for composites produced by 6- and 8-year-old or 8- and 10-year-old children.

In Section 8.3.3 it was explained that these findings were in contrast to the findings of Davies *et al.* (1989) (who found no significant difference between likeness ratings for composites constructed by 6- to 7- and 10- to 11-year olds using the Photo-FIT system) and Schwartz-Kenney, Norton, *et al.* (1996) (who found a statistically significant difference between likeness ratings for composites constructed by 5- to 6 and 8- to 9-year-olds using the Identikit system). However, when considering the actual likeness scores composites created by children received comparisons can be made with the research studies described in Appendix Va. The findings of the present study approximately matched the scores reported by Davies *et al.* and were only slightly lower than the scores reported by Schwartz-Kenney, Norton *et al.* Furthermore, the corresponding likeness scores of the present research were higher than the scores

reported by Memon using the E-FIT system with 7- to 11-years-olds (A.Memon, personal communication, 14th May 2004).

For the objective measure of composites provided by the matching task, over 66% of composites were correctly matched to target by at least one participant. Again the composites produced by adults were of a statistically significantly better quality than the composites created by all ages of children. However, there was no statistically significant difference between the number of correct matches for composites produced by 6-, 8- or 10-year-old children. Section 8.4.3 described how this finding was consistent with research by Flin *et al.* (1989) who found a statistically significant difference between adults and children's Photo-FITs but did not find a statistically significant difference between 8- to 9-year-olds and 11- to 12-year-olds (although this difference approached significance).

In terms of the actual scores of the matching task, 39% of participants correctly matched composites constructed by adults, consistent with research conducted using the E-FIT system by Turner (2004). The results of the correct matches for composites constructed by children were compared to the only published study which has examined the utility of children's composites (Flin *et al.*, 1989). The percentage of correct matches reported by Flin *et al.* were higher than the averages in the present research. Additionally, Flin *et al.* found that on average, composites created by participants of all ages were correctly sorted above chance. In contrast, the present study found that composites created by 6-year-olds, and composites constructed by 8-year olds using the visual prompts were being matched, on average, below chance level.

In Sections 8.3.3 and 8.4.3 it was argued that the different rating and matching findings in the present research described above may be due to the different circumstances composites were evaluated under. For example, the different likeness rating scales used by Schwartz-Kenney, Norton, *et al.* and Davies *et al.* and the six person line-up used in the objective matching task by Flin *et al.* It was also argued that differences may also be due to the varying circumstances the composites were constructed under. For example, the inclusion of artistic enchantment and the non-inclusion of a delay or the younger age group of children in Flin *et al.*'s study.

A possible explanation for the significant findings between 6- and 10-year-olds children for the likeness scores but not for utility scores may have been due to the targets used. As discussed in Section 8.4.3, the identification of a number of consistently 'incorrectly' selected targets may be due to the fact that they were 'type likenesses' of each other (in that they shared common characteristics). As described in Chapter 4, a facial composite is intended to be a 'type likeness' rather than a photographic representation. A possible interpretation is that a number of the E-FIT composites which were incorrectly matched in the Study 4 may have been good 'type likenesses' (rather than specific likenesses) which would have prompted a number of possible matches, which is all that is required of an E-FIT. This issue will be considered in more detail in relation to the identification task following composite construction below and in terms of methodological limitations in Section 9.3.

Variations within the age groups for likeness and utility scores have been illustrated. However, although the results showed a general improvement in the likeness and the utility of the composites produced with age, composites constructed by adults and older children were not always of better quality than those created by younger children: as

illustrated in Sections 8.3.3 and 8.4.3 younger children's composites were included amongst some of the best composites (consistent with findings described by Flin *et al.*, 1989). Indeed, the only composite which was correctly matched by all participants was produced by an 8-year-old. In response to the questionnaire survey, the majority of operators appeared to focus on a single age at which children may be interviewed and some operators reported refusing to interview children simply on the basis of their age. The youngest age of child the majority of operators had interviewed was 12-years and most operators stated the youngest age child they would be willing to interview was 10-years. The results from the evaluations of composites indicate that if operators are willing to interview 10-year-old child witnesses then there is no reason to automatically discount younger children (as young as 6-years) from constructing composites simply on the basis of their age.

As described above, adults and older children would have spent longer and made further changes if more time had been provided, whereas, very few 6- and 8-year olds wanted to make further changes, and if they did they only specified an average of one change. Furthermore, the holistic system E-FIT also allows witnesses to make configural changes to a composite. Children and adults may use both kinds of facial information (featural and configural) depending on the requirements of the task and its degree of difficulty. Therefore, working on a feature in E-FIT, although shown within a whole face, may disrupt configural processing abilities which are weaker in young children (especially 6-years of age and under) compared with older children and adults (e.g. Friere and Lee, 2001; Mondloch *et al.*, 2002). Adults and older children may be able to utilise the holistic aspect of E-FIT more than younger children (although there was evidence that younger children did use some holistic processing) and switch between the two kinds of facial information. They may also want to go on to make more

configural changes than younger children, resulting in better quality composites. Again it was argued that these findings provide support for a quantitative change in the development facial processing described in Section 3.2.2 i.e. younger children did not want to make as many changes to the composite they were presented with.

The positive effects of the visual prompts on children's descriptions in Study 3 were not extended to children's production of facial composites in Study 4. In Sections 8.3.3 and 8.4.3 it was proposed that this was possibly due to the small effect sizes reported for Prompt in Study 3. A previous study by Flin *et al.*, 1989 asked judges to sort children's initial verbal descriptions (in addition to sorting the facial composites constructed using the Photo-FIT system). Results showed that initial verbal descriptions were significantly superior to the Photo-FITs in that they provided a better guide to likeness at each age group tested. This result has also been reported for an adult sample (Christie *et al.*, 1981). Perhaps greater use could be made of the children's selection of the visual prompts, supporting previous suggestions that greater use be made of descriptions in the forensic context (e.g. Christie and Ellis, 1981; Flin *et al.*, 1989).

In Section 4.3.2 it was described how Davies *et al.* (1989) suggested that the possible areas in which children may be lacking most in composite construction are either the retrieval of specific facial details or in the translation of such details into verbal terms, or both. Studies 2 and 3 demonstrated that with appropriate prompting, children can provide useable facial descriptions for composite construction. This perhaps suggests that it is unlikely that the difficulty for children lies in the retrieval of specific facial details. Furthermore, the prompts were designed to map directly onto the composite system (as described in Section 7.1.4) therefore, it is unlikely that the difficulty for children lies in translating the information contained in the verbal descriptions into an

accurate visual image again - at least in terms of translating children's initial facial descriptions to an initial composite image.

As illustrated in Figure 4.2 (Chapter 4) the present studies specifically focused on assisting children with verbally communicating a description of a visual impression of a target to a composite operator, and assisting an operator in transforming this verbal description back into a visual image by helping them to input the description into the E-FIT system. Perhaps the age differences observed in the present study are due to problems translating the information contained in the verbal descriptions into an accurate visual image *following* the presentation of the initial composite face. For example, although older children and adults went on to specify further changes to be made to their E-FIT composites if more time had been provided an alternative explanation to the one provided previously was based on children's language development (as described in Section 2.1.3). Maybe the younger children could not verbalise further changes, or they did not understand the question. Perhaps an appropriate interview technique similar to the prompts used during the interview process could be used with children whilst constructing a composite. Further work would be needed to substantiate this interpretation.

An identification task was included following composite construction in order to determine whether participants had attended to the target and there appeared to be a group of children who had simply failed to attend to the target (55% of 6-year-olds made an incorrect identification). It was argued in Section 8.2.3 that it was possible these children were simply not encoding the target face and subsequently could not recognise it. Children may have simply encoded the face at a level which enabled them to produce a 'type likeness', which was enough to construct a composite but not enough

to correctly identify the target in an identification task. This would also explain the low numbers of correct matches in the matching evaluation task (Experiment 7) for composites constructed by children compared to adults (where participants were asked to match composites to a photograph within an array of type likenesses). Alternatively, the identification task may not be telling us whether children encoded the face or not as there was no statistically significant effect of identification on the quality of the composites which were constructed (as measured by ranking, rating and matching tasks).

In summary, Study 4 provides an important theoretical contribution to informing research on how children can be assisted through developmentally appropriate techniques to produce composites of unfamiliar faces. This is important because, as described in Section 4.3.2, in comparison to the number of studies conducted with adults, there are far fewer studies which have examined children's abilities to construct facial composites. Furthermore, prior to this thesis there had been no examination of children's abilities to construct composite using the computerised composite system E-FIT. Finally, the research findings provide some further support for the most recent research which suggests a quantitative (rather than a qualitative) shift in children's facial processing.

To summarise the research findings, children of all ages were able to describe and construct a composite of an unfamiliar target using a set of visual or verbal prompts and the E-FIT system. Results showed a consistent increase with age. This is as expected because both tasks involve a larger element of recall than recognition which produces the largest increments in performance with age (Flin *et al.*, 1989). As discussed, this consistent increase provides support for a quantitative rather than a qualitative shift in

children's performance with age and as Davies *et al.* (1989) commented, should a child's evidence be discounted because s/he provides one fewer descriptor than their older counterparts? Finally, throughout the findings from studies 2, 3 and 4 there has been a theme of variability amongst children's performance and it was argued that there is no reason to discount younger children from providing descriptions and constructing composites simply on the basis of their age. As will be described in Section 9.4 children's abilities as witnesses depend on a number of interacting factors including the individual characteristics of the child. Age is only one factor in the determination of children's performance and other factors may provide better predictors of children's performance.

9.3 METHODOLOGICAL CRITIQUE

There are a number of methodological factors that may have influenced the findings of the experimental studies reported here, and these are now considered in turn.

Foremost, it is appropriate to comment on some of the implications of conducting research with children in a school setting. Collecting the data for this thesis worked well in the classroom setting: the tasks were well suited to the school environment as they could be set up in any available space; the school was a familiar environment; children accepted the tasks and were happy to participate; ethical requirements could be met; and guidelines for interviewing children (Home Office, 2002) were followed as far as possible. However, due to the artificial nature of many aspects of this experimental environment, as with any research of this kind, the methodology of the thesis may be questioned on the grounds of its ecological validity.

The studies did not recreate the stressful conditions typical of many criminal encounters as all participants were volunteers rather than real, potentially anxious or traumatised victims or witnesses. Additionally, there was no stress at retrieval: there was no real consequentiality, as the participants knew that they were simply participating in an experiment, whereas in an actual police investigations the witness is attempting to bring an offender to justice.

Due to restrictions imposed by the schools used in the present study audio recording was used instead of video, supplemented by notes made by the author throughout each interview. A more detailed examination of children's performance at providing facial descriptions and constructing composites could have been provided by video recording interviews during studies 2, 3 and 4. For example, non-verbal descriptions offered by children (particularly in Study 2) have been examined in more detail. If video recording of the composite interviews during Experiment 5 were possible it would have allowed a more detailed examination of children's composite construction behaviour. ABE (Home Office, 2002) also stresses that interviews with child witnesses should be video recorded and following this guidance could make the interview process more valid, forensically. This is of particular importance when using prompts with children (in Studies 3 and 4). Children may not verbalise their choice of prompt and a video would illustrate the child's selection and could not be open to scrutiny if presented to a court as evidence.

Again, due to the school setting, in Studies 3 and 4 children viewed videos of target faces rather than real faces in a live encounter. The use of a video as the facial stimuli enabled control of a participant's exposure to a target, in terms of pose and so forth, and was appropriate for the school setting. It also focused participants' attention to the target face (as illustrated by the free description data in Study 3). However, this may be rather different to the representation a real witness would have for the face of a suspect.

Research has shown that the circumstances under which the target is observed are important. For example, paraphernalia affects children's judgement of identity (e.g. Friere and Lee, 2001). Therefore, the quality of a witness' recall is likely to be different to when the face is viewed under optimal laboratory conditions. Although the use of a video of facial stimuli is more valid than a photographic facial stimuli, the most ecologically valid technique would have been a 'live' exposure to a target prior to constructing a composite (e.g. as used by Davies *et al.*, 1989; Schwartz-Kenney, Norton, *et al.*, 1996).

A one-day delay between participants being exposed to the target and recall was incorporated in Studies 3 and 4 to add ecological validity. In actual police investigations the delay between the witnessed event and the construction of a composite may range from two days to six months, although the 'typical' delay experienced by the Metropolitan police has been estimated at around one week (A. Parry, personal communication, 14Th November, 2001). Due to schools' timetable restrictions one day was the maximum delay which could be employed. A longer delay may have provided more dramatic age differences as research investigating retention and forgetting has indicated that although children's memories of events over time can be quite good, young children's performances often decrease over time, more than older children's and adults' performances (Brainerd and Ornstein, 1991). However, not all children who give descriptions of strangers must do so after long delays. Schwartz-Kenney, Bottoms and Goodman (1996) stated that children who must recall the identity of strangers in a legally relevant context may have to make reports weeks or months after the original event. However, the Metropolitan police do not tend to interview after such long delays unless the crime witnessed was serious (A. Parry, personal communication, 14Th November, 2001). Finally, although studies have shown that verbal descriptions

deteriorate over time (Ellis *et al.* 1980), no decrease in the perceived quality of composites over time has been shown (Davies *et al.*, 1978). In retrospect the questionnaire survey of composite operators could have obtained information regarding the average length of time between children witnessing an event and constructing a composite.

The present studies with child participants incorporated a practice interview (Poole and Lamb, 1998) in order to establish the ground rules and to allow children to practice indicating that they do not understand or do not know the answer to a question (e.g. Mulder and Vrij, 1996). The inclusion of such task was supported. The discussion of the results of Experiment 2 highlighted the possible effect of the practice task on the content of children's descriptions, in particular children's inclusion of configural information in their free descriptions following the practice task. Although, as the practice task was not included as an independent variable effect (as the positive effects of practice interviews have already been established (e.g. Poole and Lamb, 1998)) the possible benefits of the use of houses as practice stimuli could not be assessed. As described in the discussion of Study 3 and Study 4 (Sections 7.4.1 and 8.2.3 respectively), children's use of the "don't know" prompts provided support for the inclusion of a practice task in the use of these prompts (Schwartz-Kenney, Norton, *et al.*, 1996).

Additionally, in the present study only Caucasian male targets were used⁹³. Although not a serious confound, a child who has witnessed an actual event could not select the suspect they would have to construct a composite of. Future research in this area should certainly address the issue of race, particularly cross-race composite production, by

⁹³ Given the lack of existing research with children's descriptions of faces and composite constructions it was decided to use the Caucasian male database as it contained the largest number of exemplars.

using a more diverse range of targets, together with the relevant feature databases. It would also be useful to explore the effect of race on the content of children's descriptions.

As described in Section 8.1.2 Study 4 employed an experimental design for composite construction in order to allow a large scale study to be conducted as an initial first step in what may potentially be a whole series with children. As such a number of restrictions were included such as participants having a limited time in which to construct their composites. As described, this time limit was unrealistic compared to the time operators spend interviewing a child to construct an E-FIT and therefore may have masked or exacerbated certain findings related to age and prompt. Additionally, the operator did not use a specialised graphics package with which to refine their likeness and as previously described in Section 8.5.2 artist elaboration can have an impact on overall image quality (Gibling and Bennett, 1994). Operator effects are also an important consideration and the existing research indicates that the quality of a composite can be influenced by the skill and the experience of the operator (e.g. A. Memon, personal communication, 14th May 2004; Davies *et al.*, 1983; Koehn and Fisher, 1997). The operator in the present study only had limited training in the use of the E-FIT system. Therefore potential practice effects were controlled for by counterbalancing the order of the age of participant interviewed.

The evaluation tasks used in Study 4 provided different measures of the quality of the composites produced. The ranking and rating tasks provided a subjective likeness of the composites, requiring participants to make comparative judgements. The matching task provided a measure of the objective utility of the composites, requiring participants to make an absolute judgement. Although measuring the objective utility of a composite is

the more ‘realistic’, or ecologically valid, of these measures, subjective likeness scores provide additional information to the utility measure. For example, the utility (matching) task provided information regarding whether or not a composite is identifiable whereas the ranking and rating likeness tasks provided information about whether a composite was a poor or a good likeness. Indeed, the data from Study 4 showed some differences between the ranking, rating and matching evaluation measures (although the ranking and rating data did positively correlate with the matching data). The fact that they both measured the utility of the composites in different ways was the rationale behind using both measures. However, in the matching task participants were not given the option of a “no match” response. The inclusion of this option would have identified the composites which had the *least* utility as well as those which had the most. One implication for further research using array matching tasks as a measure of composite utility is that the instructions should perhaps be that the target “may or may not be present”⁹⁴ as is the case with police line-up procedures.

The most ecologically valid task to determine the quality of the composites constructed would have been to use a naming task as described in Section 4.3.1, where judges familiar to the targets are used to evaluate the quality of the composites by providing names for any composite they recognise. Such a task is the most analogous to the use of composites in actual police cases where someone who is familiar with the person portrayed in the composite and will be able to identify them to the police. This approach has been used by Brace, Pike, Kemp, Turner and Bennett (in press) and been shown to provide ecologically valid data. However, as with the matching task, such a measure does not include information about relative differences between identifiable composites as the ranking and rating tasks do. Therefore, an important implication for research into

⁹⁴ In the present study, the participants were aware that the array was always target-present.

composite construction is that multiple evaluation measures should be employed when evaluating the quality of composites. In addition, participants completed the evaluation tasks alone (without the presence of the experimenter). If participants had completed the tasks with the experimenter present they may have employed different criteria for their decisions. For example, participants may have been less likely to make a guess if the experimenter was recording their responses.

Although a 'typical' interview for constructing a facial composite image with adults was described in Chapter 4, when asked in the questionnaire survey, no operators described the same, or even a similar interview procedure to each other. Therefore, there was not enough information to determine a typical interview structure which could be included as a control condition throughout the thesis, and in particular with child participants in Experiment 5.

Finally, in Experiments 2, 4 and 5, children were withdrawn individually from class, performed the task and then returned to be replaced by another participant. Although children were requested not to talk about the task there remains the possibility of diffusion of information amongst participants due to the novelty of the task, anxiety, or a fear of failure.

In conclusion the methodology employed was conducive to the situation (the school setting, the time scale of the thesis and so on). The majority of the issues raised are inherent in much forensically applied research, as there are obvious ethical barriers to experimentally mimicking the experience of being a real victim or witness of crime. However, the results must be treated with a certain amount of caution and it is not possible to be certain that the same results would be found with children who have been

either the witness to or the victim of an actual crime. Research incorporating longer delays and emotionally significant ‘live’ interactions is required. Such future research along with other suggestions is discussed below.

9.4 FUTURE DIRECTIONS FOR RESEARCH

Although the results from this study are encouraging, they indicate the need for more research on this topic. Future directions for research relevant to children’s composite production are considered below.

Following the initial study of children’s abilities to construct composites with the computerised composite system E-FIT in this thesis, future studies could explore children’s abilities to construct E-FIT composites under optimum conditions, in the presence of a target or with an experienced E-FIT operator for example. As described in Section 4.1.3, studies have found significant target present/target absent effects with computerised composite systems, emphasising that these systems are able to produce more accurate likenesses than the older, non-computerised systems (Davies *et al.*, 2000; Koehn and Fisher, 1997; Kovera *et al.*, 1997). Davies *et al.* reported that judges achieved matching accuracy scores of nearly 88% for E-FITs constructed under target-present conditions. A study including such an optimum condition would allow a more detailed investigation of children’s abilities to construct composites with the E-FIT system and determine whether any differences in composite quality observed are due to composite system effects, witness effects, or perhaps a combination of both.

The studies in the present thesis focused on children’s initial facial descriptions. As illustrated in Figure 4.2 (Chapter 4), the studies specifically focused on assisting children with verbally communicating a description of a visual impression of a target to

a composite operator, and assisting an operator in transforming this verbal description back into a visual image by helping them to input the description into the E-FIT system. As suggested in Section 9.2.3 above, future research could investigate how children compare their perception of the initial visual likeness they are presented with to their memory for the target face. For example, by examining children's abilities to identify and verbalise what changes should be made to a composite in order to more closely resemble the face they are trying to re-create. At the end of the composite task in Experiment 5 many participants still wanted to make changes to the E-FIT, particularly the older children and adults. It is possible that younger children could not verbalise further changes and an appropriate interview technique similar to the prompts used during the interview process could be used to assist children during composite construction.

As discussed in Section 8.5.2 a number of measures were briefly noted by the operator throughout the composite interview with children. For example, whether participants changed any of their initial descriptions or the number of exemplars of a feature a participant looked at. A study with the main purpose of examining children's composite construction could examine the changes children make to the initial E-FIT face shown. For example: how many, and which, features are changed, resized, moved; the number of exemplars looked at for each feature. The research studies cited in Section 3.2.2 show that although 6-year-old children use both featural and configural information to process faces, it is not until 10-years of age that they have become increasingly more reliant on using configural information to distinguish between faces (e.g. Mondloch *et al.*, 2002). Further studies could examine the types of changes made to the initial E-FIT face children are presented with, in particular in terms of whether children make featural

changes or configural changes, how this compares to adults, and the effect this has on the evaluation of a composite.

As described in Section 2.3.3, it has been repeatedly demonstrated that asking children to draw during verbal interviews can also facilitate children's verbal recall (Gross and Hayne, 1999; Strange *et al.*, 2003). Indeed in Study 2 many children used non-verbal descriptions which included drawing features in the air or on the table with their hand and some children asked if they could draw on paper. The most frequently used material with child witnesses reported by operators in Study 1 was a pen and paper. Furthermore, research has shown that drawing may decrease the risk of error introduced by the interviewer in the form of distractor props, in that it permits the child to create his or her own external cues (Brennan and Fisher, 1998). However, another consideration when interviewing children is that drawing extends the duration of the interview relative to a verbal interview – time was the main concern of the operators. In terms of concentration, allowing a child to draw may focus them on the task. Perhaps the use of drawing in combination with the prompts could potentially be used to enhance reports of facial descriptions provided by children.

Throughout all of the findings in the present thesis there has been a theme of variability amongst participants' performance. Despite the group effects in the studies, there were differences not only between but also within age groups in the children's performance, suggesting that age is not always the best predictor of children's competence on such tasks and that children's abilities as witnesses depend on a number of interacting factors including the individual characteristics of the child. Perhaps other person variables, such as verbal ability and intelligence may also be related to the quality or quantity of facial descriptions. It may be useful for operators to know if particular interview strategies are

helpful for children with particular characteristic styles or specific difficulties. Bruck and Ceci (1999) stated that we are a long way from understanding the course of individual differences although recent research has begun to examine this.

For example, Willcock and Hayne (2003) indicated that a drawing technique which facilitated the recall of all children was of particular benefit when interviewing children with low IQ and memory skills. Brown and Pipe (2003) examined whether those children who were most likely to benefit from the NE or alternative technique could be predicted. A short form of the Wechsler Intelligence scale for children and measures of meta memory, narrative ability and socio-economic status were taken. Brown and Pipe found that measures of intelligence were predictive of the amount recalled for children who were shown the NE cue cards but not for other children (who received verbal prompting). Finally, Elischberger and Roebbers (2001) measured verbal intelligence in their study using non specific verbal prompts. Results showed a moderate positive effect between verbal intelligence and memory quantity performance especially for the 8-year-old children and that children from both the higher and lower verbal intelligence groups were able to use the verbal prompts to an almost equal extent. Although interviewers benefit from general guidance about child development in practice they have to deal with individual children. Therefore, an important next step in this area is to conduct research that will assist interviewers to adjust their practice to suit the particular strengths and deficits of the individual child. This will ultimately bring operators a step closer to providing an optimal environment for children to provide descriptions and construct composites of unfamiliar faces.

If findings are to be most applicable to the forensic contexts that initially prompted such research, future modifications must include efforts to simulate the varied contexts of

real crime investigations. As described above, in order to increase the ecological validity of these studies there is a clear need to supplement these findings with naturalistic studies including live interactions (e.g. as Davies *et al.*, 1989 and Schwartz-Kenney, Norton *et al.*, 1996) and incorporating longer delays. Although not all children who must give descriptions of strangers in legally relevant contexts give their reports after long delays or about emotionally troubling events, some certainly do. Thus research designs incorporating longer delays, emotionally significant interactions with strangers, and so on are indicated.

The results from the existing experimental studies are encouraging with regard to the new and original prompts developed. The next step is to extend these research findings to examine the use of the visual prompts by practitioners with actual child witnesses who are recalling under stressful conditions. As Clifford (2002) stated, almost no research has been conducted with forensic child witnesses. Therefore, interviewing and observing of police operators using the visual prompts with actual child witnesses who are recalling under difficult conditions would enable an exploration of the utility of visual prompts, and identify ways in which they need to be refined in order to be most valuable to composite operators and to child witnesses.

Finally, there are some current developments in composite construction systems such as Eigen-FIT and Evo-FIT which were introduced in Section 4.1.2. These new generation of systems are currently in the primary stages of evaluation with adult participants therefore an important future direction for research relevant to children's composite production will be an investigation of children's abilities to use these systems.

9.5 IMPLICATIONS

9.5.1 IMPLICATIONS FOR METHODOLOGY

The present thesis has some general methodological implications for future research. First, as described above, subjective and objective evaluations of composites provide different measures of quality and as such further research should include more than one evaluation measure. Secondly, both by-item and by-participant analyses should be employed when analysing evaluation measures of composites data to ensure both participants and composites are treated as random factors. Ideally analyses should also be combined to ensure results can be generalised fully (as in the analyses of the results of Experiments 6 and 7). Finally, future witness research with children should include at least three age groups (e.g. 6-, 8- and 10-years). Previous research has not included the younger age group of 6-year-olds (e.g. Flin *et al.*, 1989) or the middle age group of 8-year-olds (e.g., Davies *et al.*, 1989) and in some cases the present research reported differences between 6- and 10-year-olds but not between 6- and 8-year-olds and 8- and 10-year-olds.

9.5.2 IMPLICATIONS FOR POLICING POLICY AND PRACTICE

A number of findings discussed within this thesis have potential implications and applications for policing policy and practice.

The encouraging results from the experimental studies described above with regard to the prompts developed may be used to develop techniques to assist police composite operators in obtaining verbal descriptions and facial composites from child victims and witnesses. Younger children have rarely had access to this level of justice, often restricted to minimal interviews providing little or no descriptive detail of offenders in very serious cases (C. Clark, personal communication, 18th June 2004). This is important given continuing legislation aimed at facilitating the reception of children's

testimony (e.g. the Youth Justice and Criminal Evidence Act, 1999). The findings from this research with children may also lead to the development of the prompts for use with other vulnerable witnesses, including elderly witnesses and adults with learning disabilities (as specified in ABE, Home Office, 2002).

The impact of the research may contribute to existing discussions with Aspley Ltd. (who produce and distribute E-FIT) with the possibility that the prompts could be included within the E-FIT program.

Of interest to practitioners will be some kind of cost-benefit analysis: will the use of the prompts cost any more in terms of time, training, or equipment needed? In terms of time the results from Study 3 and Study 4 indicated that the visual prompts (and indeed the verbal prompts) reduced the time spent interviewing children compared to operators responses in Study 1. This would suggest that no increase in time cost would be expected as a result of using any of the prompts as an interview procedure. In terms of the amount of training required, this would be minimal. The prompts do not suffer from the practical limitations of other non-verbal retrieval aids such as extensive training of the interviewer and/or special preparation/training procedures before the actual interview for the child witness (as with the CI or NE procedures). Finally, in terms of equipment needed the amount and cost of the equipment would also be minimal, particularly if the prompts were presented as part of the E-FIT program as previously described.

Operators' responses to the questionnaire survey highlighted the absence of a training course specifically to provide guidance on constructing composites with children, and a number of operators stated that they would like to receive some training for constructing

composites with children. Providing a training course, specifically on producing composites with children, is an important area of future policy and practice development. However, what the training should comprise, how it should be delivered and how to sustain skills and knowledge obtained in training would need to be identified.

As an aside, from the questionnaire responses, speaking to operators at conferences and informal interviews with operators the issue of operators feeling they are an isolated group was frequently raised. Operators were not mentioned at all in the MoGP (Home Office, 1992). An important first step has been made as operators are mentioned in ABE (Home Office, 2002). However, this guidance only includes one paragraph relating to operators. As a result, many operators do not consider themselves as 'interviewers' and therefore do not follow the government's guidance. The introduction of more full time operators (which was suggested by a number of operators in the questionnaire) may lead to a more established and experienced group which would in turn address some of the main issues with child witnesses. Indeed, a questionnaire of operators since publication of ABE (Home Office, 2002) may be useful to determine exactly how many operators consider themselves as 'interviewers' rather than just operators.

9.6 CONCLUDING COMMENTS

The present thesis sought to gain a detailed understanding of children's verbal descriptions of an unfamiliar face, and how children might be assisted, through appropriate interview techniques, to provide computerised facial composite constructions of an unfamiliar face. Prior to this thesis there had been no explicit attempt to combine developmental, eyewitness, facial recognition, facial recall research to the composite construction interview process with children. Therefore, the thesis has

a number of important theoretical contributions. These include showing how more detailed examinations of the content of children's descriptions can contribute to a more detailed theoretical account of how children recall and describe unfamiliar faces. This knowledge can be applied to provide developmentally appropriate techniques to describe unfamiliar faces and construct facial composites. Furthermore, the findings from the present studies may be used to inform one of the long standing debates in the literature concerning whether, and to what extent, adults and children differ in the way in which they process information about unfamiliar faces. The findings provide some support for the most recent research which suggests a quantitative (rather than a qualitative) shift in children's facial processing (see a review by Gilchrist and McKone, 2003). Finally, the results suggest that restricting research to looking at children's performance across age groups may under-emphasise variation in children's abilities. In summary, all of these theoretical contributions have important practical implications for the ways in which children are interviewed.

Within the thesis, three research questions were addressed. In response to these questions the findings from the present thesis demonstrate that: (1) children from the age of 6-years are able to effectively describe unfamiliar faces when prompted (in terms of the content and quantity of descriptions useful for subsequent composite construction); (2) the use of a set of visual prompts is an appropriate interview technique which can be used to significantly enhance children's unfamiliar facial descriptions, in terms of reducing time taken, increasing the number and accuracy of children's descriptions; (3) children from the age of 6-years can produce facial composites of an unfamiliar face using a set of visual or verbal prompts and the computerised composite system E-FIT under experimental conditions. Finally, there were large variations within age groups in all studies, indicating that children should not

be excluded from providing facial descriptions and constructing facial composites simply on the basis of their age.

In conclusion, the thesis is an initial step to contributing to the research in this area and future research needs to examine the later stages of the composite construction process with children.